Regulation of Unpredictable Effects of Decision Making Systems is Non-trivial

Jussi Karlgren

Recent and coming technical advances are delegating decision making in new arenas of human activity to information systems. The main research direction in artificial intelligence today is machine learning, which at its most abstract level is a classification mechanism: modern information processing mechanisms are able to ingest vast and vastly growing streams of noisy and heterogenous data and turn them into decisions: stop or go, sell or buy, yes or no, safe or risky, and so forth. We can expect classification tools to move from laboratories to many more application fields in the near future. Previous waves of automation has focussed on tasks which are manual and repetitive: industrial robots have rapidly become a productive tool in manufacturing and extractive mining. Coming automation is expected to take on tasks which are routine, but not necessarily repetitive and not necessarily manual. This includes tasks which have direct impact on human lives, involving decisions about medical treatment, contractual obligations, and sentencing in criminal courts as well as on-the-spot decisions in vehicle control, security systems, and financial trading.

Managing technology-induced change and its effects through legislative systems in order to encourage and support behaviour and activities which are desirable and beneficial to the public good and dissuade from such which is not, is a careful and delicate craft. In general, legislation to cover new technical advances will be based on existing technology and existing practice. This may seen reasonable basis to build from and adds legitimacy to regulation and its application, but regulation of technology too often stumbles at the balancing line between understanding and promoting future change productively and protecting past practice and existing business models. Regulatory models that are under strain are e.g. how to understand intellectual property rights when scarcity no longer is a factor, when distribution costs are near nil, and original and copy cannot be distinguished; or how to understand editorial roles and responsibilities of organisations in the business of information dissemination when content is produced by consumers, aggregation and selection is uncoupled from production, and editorial oversight is minimal or non-existent.

1 Changes

The introduction of technology into situations changes things. If those changes are noticeable this will make people worry and react variously. New technology often has effects that are different from those originally intended, in matters both momentous and trivial. Mobile phones have had unexpected side effects on meeting planning, number memorisation, location based services, and grain price dispersion in third world countries; map and route planning technology has changed navigational behaviour and addressing practices; camera phones have made postcards obsolete.

Disruptive side effects are especially true for artificial intelligence and related technologies, designed to make decisions in situations where humans struggle to do so consistently and tirelessly. These sorts of technologies are based on learning the basis for classification, decision, and action from previously known data and processing them appropriately. This will shift initiative and power to those with access to data and processing tools from those in charge of routines

and processes today. Data intensive decision making is by its nature oligopolic: initial cost of investment is comparatively high compared to marginal cost of adding data, and there are attendant immediate rewards from scale.

Legal systems are already in place to handle the risks of oligopoly. Some of them are relevant to this application area: how they should be amended is a legal challenge.

2 Pattern Analysis

Data analytics and machine learning is not about singular data points. This notion is important to understand. The data points themselves may be of little interest and value in themselves and only valuable inasmuch they contribute to a larger pattern. Removing some data points from a data set or a data stream are not likely to change much in the overall insights gained from the analysis; inspecting those data points in isolation will not yield the insights the entire pattern would have. The patterns in past data enables analyses to infer information even from noisy future data and from future data with missing data points. This has consequencecs for the explainability of the data, the insights learned from them, and decisions made on those insights.

3 Any Classification System, Whether Human or Automatic, will Risk Errors

This may be because the experience of the system is insufficient (too little training data, or training data which lack examples of crucial states of the world), because the system is inconsistent (some other confounding, possibly unrelated, variable perturbs the application of the relevant experience), because the system applies a faulty, badly implented, or irrelevant classification algorithm, because the categories under consideration are unsuitable to the task they are being applied to, or (most typically) because the categorisation scheme is less clearcut at time of application than it seemed at time of design.

We can expect that systems built to make independent decisions will make such decisions in ways which differ from humans, frequently in interesting ways. This will mostly be to the benefit of all, but in some cases will require conventions to handle disappointment and distress.

When errors happen, their causes should be traceable. Tracing what has occasioned some decision to be made is a complex task. Increasingly, data-intensive systems which incorporate recent artificial intelligence technology, are built without an explicit representation of input data, context, or reasoning. This will make such querying, and thus, learning from errors, more difficult. In light of the fact that human cognitive processing seems to be slower than human decision making, a persuasive argument has been given that human consciousness is a mechanism not to make decisions, but to explain decisions—not least to oneself—after the fact.

Understanding where accountability and potential liability lies and tracing the decision of a working system back to its many multiple sources will involve more than technology. It is technically possible to provide computational systems with capabilities to answer questions of how a decision came about, but there are several layers of complexity and attendant layers of responsibility here. Moving a computational process into practical usage involves numerous non-trivial steps by nonoverlapping sets of system developers and other related professions. From the first steps of identifying a computational challenge, to creating an algorithm, implementing that algorithm in some computational framework, testing and veryifing the working on that algorithm on some test set, identifying and specifying what types of data the system is designed to handle, providing the data with defaults where coverage is patchy and constraints where choices are overly broad, deploying an implemented system in some system environment, identifying sources and aggregating data for the runtime system operation, training operators to use the system, formulating best practice for understanding the results, updating the system periodically to accommodate technical advance potentially contributed by non-related sources, handling potential discrepancies with previously established routines—all of these and many other activities impinge on the results of operating the system in some field of application.

4 Tracing

Tracing the origins of a decision is important even when no direct errors have happened. The argument that a decision making system needs oversight to stave off errors is easy to make in face of public worry about consequences, as delineated above, but when consequential and non-trivial decisions are made, people whose lives and activities are impacted by those decisions will for various reasons need to be able to query the system about *how* and *why* those decisions are made, even when no error has happened.

This is necessary to ensure the legitimacy of decisions, and to allow subjects to contest a decision and argue their case if they would wish that a decision were modified, to correct or remove faulty data points, to add new relevant data, to adjust inferences, to increase the understanding of how information about a subject may cause decisions to made for future reference.

Legal systems are already in place to handle accountability and legal liability. Some of them are relevant to this application area: how they should be amended to handle decision making partially through automatic mechanisms is a legal challenge.

5 Information Imbalance

Information imbalance is a major stumbling block for making such data oversight possible. Subjects of a decision may have or may be able to obtain a fairly complete set of data used to characterise them, but they will typically lack

the processing methodology and the implementations thereof used in the decision making process, and more importantly, they will lack the contextual data of other subjects in the population the decision is shaped by. Returning to the above note about how classification systems rely on patterns in data rather than individual data points, and on how patterns in the entire population are more or less similar to some individual's data, individuals cannot be expected to understand what in the data collected about them will have potential effects on the model of them the analysis can yield. The individual data points are of little or no utility to assess their worth or impact, and regulations which allow an individual to inspect, audit, and remove data about themselves of little value if no support for an overview of the data is given. That overview is only accessible to those who have recourse to the full data set and tools to process it with.

Individual data are often collected, aggregated, handled, and disseminated to other agents through platforms designed for some purpose of enjoyment or utility of individuals. The subsequent value of the individual data are typically not comprehensible at time of collection.

The downstream processing and use of those data — for purposes such as credit scores, risk assessment, employability, insurance rates, and other similar decisions — is typically done by other organisations that frequently are unaware of how the data they use came about, and frequently do not even have access to the original data themselves, since they have been repackaged by third party analytic organisations that provide them with operational insights.

Legal systems are already in place to handle auditing procedures, transparency, and ownership of information. Some of them are relevant to this application area: how they should be amended is a legal challenge.

6 Invisible Harm

The effects of automated systems may be unnoticeable in many situations. In a job seeking situation, the applicants are ranked and only one of the candidates will get the position. In a loan application those applicants who are approved for a loan are given a rate based on a credit assessments made. A quote for an insurance rate is based on a risk assessment. Researching a fare to a popular destination or lodging there reveals offers given at a certain rate, based on the attractivity of the potential customer. A educational instiution accepts a subset of those applying for admission. These sorts of assessments are today mostly done by human analysts, but if handed over to automated decision systems, the consequences are complex. Who is responsible for the decision? And more importantly, who will notice that the decision was made automatically rather than after human deliberation? And how can those decisions be audited and accountability be traced back if they turn out to be contestable?

The suspicion that such invisible harms may come about from the actions of an individual will erode trust in systems and public processes. This in turn causes worse harms to the public sphere: if the general public adopts a widespread opt-out and self-censorship strategy, this will put a damper on many future potentially valuable services and systems.

Legal systems are already in place to handle harm, loss, and damages. Some of them are relevant to this application area: how they should be amended to handle decision making where the individual is not aware of how decisions impact them and how to ensure a level of transparency to such calculations is a legal challenge.

7 The Role of Regulatory and Legislative Systems

Much of the central task for the legal profession is to ensure responsibility for consequences of human action rests with the right party and to provide regulatory mechanism to trace accountability to the right party. Traps the legal profession should avoid is to spend effort (1) to work from technical details instead of effects of technology; (2) to protect and conserve existing business models; (3) to frame regulation reactively to worrisome events; (4) to map emerging processes too closely to previously known processes. Suggested avenues for the legal profession to approach the coming challenges of automated decision making is to work to (1) ensure *auditability of information flow* to require that automated processes provide a decision trace; (2) ensure *transaction transparency* to require that individual data submitted as a partial payment for a service rendered should be made into an explicit transaction subject to approval and retraction by the individual in question; (3) ensure *explicit valuation of individual data* to require that the value of aggregated data should be made part of the audited value of a business organisation; (4) make explicit the *ownership status* of individual data, so that the release of data from one organisation to another can be traced across subsequent transactions and aggregation; (5) provide *behavioural transparency* to require automated systems with impact on individual clients, customers, or other stakeholders, to run periodical test cases and make them available for inspection in case legal procedures would benefit from them.