Big Data Society: Age of Reputation or Age of Discrimination?

by Dirk Helbing (ETH Zurich)

If we want Big Data to create societal progress, more transparency and participatory opportunities are needed to avoid discrimination and ensure that they are used in a scientifically sound, trustable, and socially beneficial way.

Have you ever "enjoyed" an extra screening at the airport because you happened to sit next to someone from a foreign country? Have you been surprised by a phone call offering a special service or product, because you visited a certain webpage? Or do you feel your browser reads your mind? Then, welcome to the world of Big Data, which mines the tons of digital traces of our daily activities such as web searches, credit card transactions, GPS mobility data, phone calls, text messages, facebook profiles, cloud storage, and more. But are you sure you are getting the best possible product, service, insurance or credit contract? I am not.

Like every technology, Big Data has some side effects. Even if you are not concerned about losing your privacy, you should be worried about one thing: discrimination. A typical application of Big Data is to distinguish different kinds of people: terrorists from normal people, good from bad insurance risks, honest tax payers from those who don't declare all income ... You may ask, isn't that a good thing? Maybe on average it is, but what if you are wrongly classified? Have you checked the information collected by the Internet about your name or gone through the list of pictures *google* stores about you? Even more scary than how much is known about you is the fact that there is quite some information in between which does not fit. So, what if you are stopped by border control, just because you have a similar name as a criminal suspect? If so, you might have been traumatized for quite some time.

Where does the problem originate? Normally, the groups of people to distinguish are overlapping -- their data points are not well separated. Therefore, mining Big Data comes with the statistical problem of false positives and false negatives [1]. That is, some people get an unintended advantage, while others suffer an unfair disadvantage -- an injustice hard to accept. Even with the overly optimistic assumption that the data mining algorithm has an accuracy of 99.9% -- when applied to 200 Million people, there are hundreds of thousands of people who will experience a wrong treatment. In medicine, the approach of mass screenings is therefore highly controversial [2]. Are you willing to sacrifice your breast or prostata for a wrongly diagnosed cancer? Probably not, but it happens more often than you think.

Similarly, tens of thousands of honest people are unintentionally mixed up with terrorists. So, how can you be sure you are getting your loan for fair conditions, and do not have to pay a higher interest rate, just because someone in your neighborhood defaulted? Can you still afford to live in an easy-going multi-cultural quarter, or do you have to move to another neighborhood to get a reasonable loan? And what about the tariff of your health insurance? Will you have to pay more, just because your neighbors do not go jogging? Will we have to put pressure on our facebook friends, colleagues, and neighbors, just to avoid possible future discrimination? And what would be the features that play out positively or negatively? How much Coke on our credit card bill will be acceptable to our health insurance? Is it ok to drink a glass of wine, or better not? What about another cup of coffee or tea? Can we still eat meat, or will we get punished for it with higher monthly rates? Would there be a right way of living at all, or would just everyone be discriminated for some behavior, while perhaps getting rewards for others? The latter is surely the case.

This might be fine, if everybody would benefit on average, but unfortunately this is rather unlikely. Some would be lucky and others would be unlucky, i.e. inequality would grow. But similar to stock markets, it would be difficult to tell before, who would benefit and who would lose. This is so not just because of the random distribution of individual properties, but also because the parameters of the data mining algorithms can be determined only with a limited accuracy. However, even tiny parameter changes may produce dramatically different results (a fact known as "sensitivity" or "butterfly effect") [3]. In other words, while the miners of Big Data may pretend to take more scientific, better and fairer decisions, the results will often have a considerable amount of arbitrariness. Many data miners probably don't know about this or don't care. But the fact that lots of algorithms produce outputs without warnings of their limitations creates a dangerous overconfidence in their results. Moreover, note that the choice of the model can be even more critical than the choice of parameters [4]. That's basically why people say: "Don't believe a statistics that you haven't produced yourself."

The problem is reminiscent of the experiences made with financial innovations. People used models without questioning their validity enough. It was discovered too late that financial innovations may have negative effects and destabilize the markets. One example is the excessive use of credit default swaps, which package risks in ways that buyers don't seem to understand anymore. The consequence of this was a financial meltdown that the public has to pay for at least for another decade or two. It is no wonder that trust in the financial system dropped dramatically, with serious economic implications (no trust means no lending). This time, we should not make the same mistakes, but rather use Big Data in a trustworthy, transparent, and beneficial way. To reap the benefits of personalized medicine, for example, we need to make sure that personal medical data will not be used to the disadvantage of patients who are willing to share their data in favor of creating a public good -- a better understanding of diseases and how to cure them.

In fact, we have worked hard to overcome discrimination of people for gender, race, religion, or sexual orientation. Should we now extend discrimination to hundreds or even thousands of variables, just because Big Data allows us to do so? Probably not! But how can we protect ourselves from such discrimination? In order to avoid that the information age becomes an age of discrimination fueled by Big Data, we need **informational justice**. This includes to establish (1) suitable quality standards like for medical drugs, (2) proper testing, and (3) fair compensation schemes. Otherwise people will quickly lose trust in Big Data. This requires us to decide what collateral damage for individuals would be considered tolerable or not. Moreover, we need to distinguish between “healthy” and “toxic” innovations, where “healthy” means innovations that produce long-term benefits for the economy and society (see Information Box). That is, the overall benefit should be bigger than the disadvantage caused by false positives, such that the corresponding individuals can be compensated for unfair treatments.

There are two fundamentally different ways to ensure a "healthy" use of Big Data and allow victims of discrimination to defend their interests. The classical approach would be to create a dedicated government agency or institution that establishes detailed regulations, in particular quality standards, certification procedures, and effective punitive schemes for violations. But there is a second approach -- one that I believe could be more effective for companies and citizens than complicated legal and executive procedures. This framework would be based on next-generation reputation systems creating feedback loops that support self-regulation.

How would such a next-generation reputation system work? The proposal is to establish a *Global Participatory Plattform* [5], i.e. a public store for models and data. It would work a bit like an *appstore*, but people and companies could upload not only *apps*. They could also upload data sets, algorithms (e.g. statistical methods, simulation models, or visualization tools), and ratings. Everybody could use these data sets for free or for a fee, and annotate user feedbacks. It would be as if we could submit not only queries to *google*, but also algorithms to determine the answers. In this way, we could better control the quality of results extracted from the data.

So, assume we would store all data collected about individuals in a data bank (for reasons of data security, a decentralized and encrypted storage would be preferable). Moreover, assume that everyone could submit algorithms to be run on these datasets. The algorithms would be able to perform certain operations within the bounds of privacy laws and other regulations. For example, they could generate aggregate information and statistics, while privacy-invasive queries violating user consent would not be executed. Moreover, if executable files of the algorithms used by insurance or other companies using Big Data would be uploaded as well, it would allow scientists and citizens to judge their statistical properties and verify that undesirable discrimination effects are below commonly accepted thresholds. This would ensure that quality standards would be met and continuously improved.

The advantages of such a transparent and participatory approach are multifold for business, science, and society alike: (1) results can be verified or falsified, thereby uncovering possible methodological issues, (2) the quality of Big Data algorithms and data will increase more quickly, (3) “healthy” innovation and economic profits will be stimulated, (4) the level of trust in the algorithms, data and conclusions will increase, and (5) an "information ecosystem" will be grown, creating an enormous amount of new business opportunities, to fully unleash the potential of Big Data.

I fully agree with the US Consumer Data Privacy Bill of Rights [6] stating that *"Trust is essential to maintaining the social and economic benefits that networked technologies bring to the United States and the rest of the world."* A report on personal data as a new asset class, published by the World Economic Forum, therefore suggests a "New Deal on Data" [7]. This includes establishing a data ecosystem that creates a balance between the interest of companies, citizens, and the state. Important elements of this would be: transparency, more control by citizens over their personal data, and the ability for individuals to participate in the value generated with their personal data.

This has implications for the design of the *Global Participatory Platform* I am proposing. Data collected about individuals would be stored in a personal data purse. Individuals could add and comment the data, have them corrected, if factually wrong, and determine, who could use them for what kind of purpose, to meet the regulations regarding privacy and self-determination. When personal data are used, both the user and the company that collected the data would earn a small amount, triggering micropayments. Finally, to keep misuse of data and malicious applications on a low level, there would be a certain reputation system, which would act like a *social immune system*.

Reputation and recommender systems are quickly spreading all over the Web. People can rate products, news, and comments. In exchange, *amazon*, *ebay*, *tripadvisor* and many other platforms offer recommendations. Such recommendations are beneficial not only for the user, who tends to get a better service, but also for a company offering the product or service, as higher reputation allows it to take a higher price [8]. However, it is not good enough to leave it to a company to decide, what recommendations we get, because then we don't know how much we are being manipulated. We want to look at the world from our own perspective, based on our own values and quality criteria. It would be terrible if everyone ended up reading the same books and listening to the same music. Therefore, it is important that recommender systems do not undermine socio-diversity.

Diversity is an important factor for innovation, social well-being, and societal resilience [9]. It deserves to be protected in the very same way as biodiversity. Modern societies need a complex interaction pattern of diverse people and ideas, not average people who all do the same things. The socio-economic misery in many countries of the world is clearly correlated with the loss of socio-economic diversity. While some level of norms and standardization appears to be favorable, too much homogeneity turns out to be bad. This also implies that we need to be careful about discriminating people who are different -- such discrimination may undermine socio-diversity.

Today's personalized recommender systems endanger socio-diversity as well. They are manipulating people's opinions and decisions, thereby imposing a certain perspective and value system on them. This can seriously undermine the “wisdom of crowds” [10], which is central to the functioning of democracies. The "wisdom of crowds" requires independent information gathering and decision-making -- a principle not sufficiently respected by most recommender systems [11].

How could we, therefore, build "pluralistic" reputation and recommender systems, which support socio-economic diversity, and are also less prone to manipulation attempts? First, one should distinguish three kinds of user feedbacks: facts (linked to information allowing to check them), advertisements (if there is a personal benefit for posting them), and opinions (all other feedbacks). Second, user feedbacks could be made in an anonymous, pseudonymous, or personally identifiable way. Third, users should be able to choose among many different reputation filters and recommender algorithms. Just imagine, we could set up the filters ourselves, share them with our friends and colleagues, modify them, and rate them. For example, we could have filters recommending us the latest news, the most controversial stories, the news that our friends are interested in, or a surprise filter. So, we could choose among a set of filters that we find most useful. Considering credibility and relevance, the filters would also put a stronger weight on information sources we trust (e.g. the opinions of friends or family members), and neglect information sources we do not want to rely on (e.g. anonymous ratings). For this, users would rate information sources as well, i.e. other raters. Therefore, spammers would quickly lose reputation and, with this, their influence on recommendations made.

In sum, the system of personal reputation filters would establish an “information ecosystem”, in which increasingly good filters will evolve by modification and selection, thereby steadily enhancing our ability to find meaningful information. Then, the pluralistic reputation values of companies and their products (e.g. insurance contracts or loan schemes) would give a pretty differentiated picture, which can also help the companies to develop better customized and more successful products.

In conclusion, I believe it's high time to create suitable institutions for the emerging Big Data Society of the 21st century. In the past, societies have created institutions such as public roads, parks, museums, libraries, schools, universities, and more. But information is a special resource: it does not become less, when shared, and it can be shared as often as we like. In fact, our culture results from what we share. At the moment, however, the world of data is highly proprietary and fragmented. It’s as if every individual owned a few words but had to pay for using all the others, and some words could not be used at all for proprietary reasons. Obviously, such a situation is not efficient and does not make sense in an age where data are increasingly important. Business and politics have pushed hard to remove barriers to the free trade of goods -- it is now time to remove the obstacles to the global use of data. Providing access to Big Data would unleash the power of information for business, politics, science and citizens. Access to Big Data is surely needed for science to provide a good service to society [12,13]. In the past, reading and writing was a privilege, which came with personal advantages. But public schools opened literacy to everyone, thereby boosting the development of modern service societies. In the very same way could we now boost the emerging digital society by promoting digital literacy and investing into transparent, secure, participatory and trustworthy information and communication systems [14]. The benefits for our societies would be huge!

References

[1] C. Chivers, How likely is the NSA PRISM program to catch a terrorist?, bayesianbiologist, see http://bayesianbiologist.com/2013/06/06/how-likely-is-the-nsa-prism-program-to-catch-a-terrorist/ and http://futurict.blogspot.ch/2013/06/why-mass-surveillance-does-not-work.html

[2] G. Gigerenzer, W. Gaissmaier, E. Kurz-Milcke, L.M. Schwartz, and S. Woloshin (2008) Helping doctors and patients make sense of health statistics, *Psychological Science in the Public Interest* **8**(2), 53-96.

[3] I. Kondor, S. Pafka, and G. Nagy (2007) Noise sensitivity of portfolio selection under various risk measures. *Journal of Banking & Finance* **31**(5), 1545-1573.

[4] T. Siegfried (2010) Odds are, it's wrong, *Science News* **177**(7), p. 26ff, see http://www.sciencenews.org/view/feature/id/57091/description/Odds_Are_Its_Wrong; J.P.A. Ioannidis (2005) Why most published research findings are false, *PLoS Medicine* **2**(8): e124.

[5] S. Buckingham Shum, K. Aberer, A. Schmidt, S. Bishop, P. Lukowicz et al. Towards a global participatory platform (2012) Democratising open data, complexity science and collective intelligence. *EPJ Special Topics* **214**, 109-152.

[6] The White House (2012) Consumer data privacy in a networked world: A framework for protecting privacy and promoting innovation in the global digital economy, see http://www.whitehouse.gov/sites/default/files/privacy-final.pdf

[7] World Economic Forum (2011) Personal Data: The Emergence of a New Asset Class, see www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf

[8] W. Przepiorka, Buyers pay for and sellers invest in a good reputation: More evidence from eBay, *The Journal of Socio-Economics* **42**, 31-42 (2013).

[9] S.E. Page (2007) *The Difference* (Princeton University Press, Princeton).

[10] J. Lorenz, H. Rauhut, F. Schweitzer, and D. Helbing (2011) How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences of the USA* **108**(28), 9020-9025.

[11] T. Zhou, Z. Kuscsik, J-G. Liu, M. Medo, J.R. Wakeling, and Y-C. Zhang (2010) Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences of the USA* **107**, 4511-4515.

[12] B.A. Huberman (2012) Big data deserve a bigger audience, *Nature* **482**, 308.

[13] F. Berman and V. Cerf (2013) Who will pay for public access of research data? *Science* **341**, 616-617.

[14] D. Helbing (2013) Economics 2.0: The natural step towards a self-regulating, participatory market society, *Evolutionary and Institutional Economics Review* **10**(1), 3-41.

Information Box: How to define quality standards for data mining

Assume that the individuals in a population of N people fall into one of two classes. Let us consider people of kind 1 “desirable” (e.g. honest citizens, good insurance risks) and people of kind 2 “undesirable” (criminals, bad insurance risks, etc.). We represent the number of people *classified* as kind 1 and 2 by N_1 and N_2 respectively. Let the rate of false positives, that is individuals who are faced with unjustified discrimination, be given by α, and the rate of false negatives be β. Then, the *actual* number of people of kind 1 is $(1-\beta)N_1+\alpha N_2$, and the *actual* number of people of kind 2 is $(1-\alpha)N_2+ \beta N_1$. Furthermore, assume that the classification is creating an advantage of $A>0$ for people classified as kind 1, but a disadvantage of $-D<0$ for people classified as kind 2. Then, each false positive classified person has a double disadvantage of $-(A+D)$, because he or she should have received the advantage A while suffering the disadvantage $-D$. This will be considered unfair and question the legitimacy of the procedure. False negatives, in contrast, those who are classified “desired” but are in fact “undesired”, enjoy a double advantage of $(A+D)$. They may also create an extra damage $-E$ to society. Overall, the classification produces a gain of $G=N_1[(1-\beta)A+\beta(A+D)]$ to individuals classified to be of kind 1 and a cost of $C=-N_2[(1-\alpha)D+\alpha(D+A)]$ to individuals classified to be of kind 2. The overall benefit to society would be $B=G-C-E$. Unfortunately, there is no guarantee that it would be positive.

To demonstrate this, let us assume a business application of Big Data, in which the economic profit P (e.g. by selling cheaper insurance contracts to people of kind 1) is a fraction f of the gain, i.e. $P=fG$. If applied to many people, the application may be profitable even if the fraction $f<1$ is quite small. Moreover, from the point of view of a company, discrimination may be rewarding even if it has an overall disadvantage to people (i.e. if the overall benefit B is negative). This is because a company typically cares about its own profits and its customers, but not everybody else. Clearly, if some insurance contracts get cheaper, others will have to be more expensive. In the end, people with high risks will not be offered insurances anymore, or only at an unaffordable price, so some victims of accidents may not be compensated at all for their damage.

Even if $\boldsymbol{B}$ is positive, the profit $\boldsymbol{P}$ may be smaller than the unjust disadvantage $\boldsymbol{U}$, which is the price that false positives have to pay. Such a business model would create a situation that I will call a **"discrimination tragedy,"** where citizens have to pay the price for economic profits, even though they are not getting a good service in exchange.

It is, therefore, in the public interest to establish binding standards for the "healthy" use of Big Data algorithms, regulating the required predictive power and the acceptable values of $\boldsymbol{\alpha}$, $\boldsymbol{D}$, $\boldsymbol{B}$ and $\boldsymbol{U}$. A cost-benefit analysis suggests to demand $\boldsymbol{B>0}$ (there is a benefit) and $\boldsymbol{B>U}$ (the benefit is high enough to compensate for unjust treatments). Moreover, $\boldsymbol{\alpha N_1}$ and $\boldsymbol{D}$ should be below some acceptable thresholds. Today, these values are often unknown, and that means we have no idea what economic and societal benefits or damages are actually created by current applications of Big Data.