Workshop Discussion Notes: Predicting Human Behavior

The Social, Cultural & Ethical Dimensions of "Big Data"
March 17, 2014 - New York, NY
http://www.datasociety.net/initiatives/2014-0317/

This document was produced based on notes that were taken during the Predicting Human Behavior workshop as part of "The Social, Cultural, and Ethical Dimensions of 'Big Data'". Not all attendees were involved in every part of the conversation, nor does this document necessarily reflect the views and beliefs of individual attendees. All workshop participants were given workshop materials prior to the event to spark discussion. This primer can be found at: http://www.datasociety.net/pubs/2014-0317/PredictingHumanBehaviorPrimer.pdf

Overview

In this workshop, participants discussed both the problems and potential opportunities posed by big data analytics when it comes to predicting human behavior. Some participants noted that while we often discuss the problems, we rarely talk about the potential benefits of predictive data both to individuals and groups. The bulk of discussion, however, was centered on the negative effects of predictive analytics, especially on unforeseen negative consequences. Several of the key issues addressed included the variable access to data; problems with data collection; the issue of re-identification; the use of prediction to shape people's futures; and the challenges associated with transparency. Some participants' conversations focused on the barriers in place when it comes to using potentially beneficial models of prediction and when such predictions have the power to harm individuals and groups, including when predictive data is useful and when it is not. The most helpful data for enhancing social services or offering humanitarian assistance may in fact be "private data," which means that access to it is limited. For example, what if movement patterns during earthquakes or other natural disasters could be predicted, allowing emergency personnel to save more lives? While this information may be open to commercial entities, the benefits may not extend to individuals. For instance, predictive data analytics may help individuals learn new things about their own behavior, allowing them to either make positive changes or to reinforce existing healthy behaviors.

Another focal point of the discussion centered on the question of, can we create legal and practical frameworks for sharing at least the aggregate data safely? Data loses its usefulness if it is cleaned too much or overly aggregated; too much data or data that has been streamlined can become meaningless. The problem of re-identification was discussed at length. Health data may benefit researchers and individuals, but there is also a risk of re-identification. How can individual patients be contacted, informed, and agree to give consent before their data is used for predictive purposes? Researchers cannot assume that everyone has the same capacity or tools to understand what the data actually reveals.

Several group members debated the utility of applying models of transparency to data-driven predictions. How can we make sure that transparency is both accessible and actionable at relevant decision points? Data sharing may take data out of context and evolving algorithms may face mission creep as data teams or organizations change over time. The predictive powers of "big data" may have tremendous effects on individuals' digital destinies. It can control the content that one sees, impacting job offers, shopping habits, and educational choices for one's children. The assumptions made by predictive algorithms may also be incorrect, which may lead to the unfair treatment of individuals and certain groups. Discrimination by algorithm is a real possibility but such discrimination may be harder to detect. On the other hand, if data is valuable for humanitarian purposes, how can we create ethical and legal frameworks that balance the benefits with the risks? How can individuals become better educated about both the benefits and risks of predictive analytics?

Themes and Discussion Topics

Opportunities versus risks

Group discussions touched on the potential opportunities created by predictive data analytics instead of just iterating the potential challenges inherent to their use. We tend to focus on the negative possibilities rather than on the benefits. Predictive analytics may lead to advances in health overall (systemic), healthcare institutions (institutional), and individuals (individual). They may help us figure out what problems we need to address in the first place, allow us to predict future problems, and ensure that we attend to the problems in the best way possible. Despite these caveats, some group members spent the majority of the allotted time discussing the negative consequences of predictive data analytics than the potential benefits.

Some participants discussed both the benefits and risks of collecting data without an intended purpose. When it came to passing Civil Rights and women's rights legislation, many citizens were indifferent to the cause. Supporters and detractors marshaled different evidence in order to tip the scale. This demonstrates that data is neither efficient nor fair in isolation; data has rhetorical weight. The future uses of predictive analytics emerged as a major concern for many participants because it's unclear how all of the data being collected today will be used down the road. Medical data, for example, may be reused in unforeseen ways. Researchers may already have access to this kind of data, but what about the data from social media? How can this data be used in the future and what controls are needed in order to make sure that it is not used for exploitation? Knowing how this data is to be used in the future, and by whom, would affect the approaches to data collection and maintenance today. Commercial entities, for example, may have less compunction about grabbing all sorts of personal data for unspecified future use. What is the data being accumulated and is it being collected simply as a function of the technology instead of being motivated by some kind of research design?

Applications of prediction

Before the group could begin talking about the implications of predicting human behavior though data analytics, participants debated what they meant by prediction and outlined its possible applications. Different variations of prediction were discussed. Breast cancer research, for instance, is only effective if you're able to pool large datasets with many variables and examples. Both individuals and groups may benefit from this sort of prediction. In some cases, prediction may also imply persuasion and manipulation. While some examples may show the need for and value of large and complex datasets, we also have to consider the impact of publication bias, or the fact that researchers tend to emphasize the results that support their own hypotheses. Prediction may mean revealing actionable information about someone in need, such as specific populations that might be hungry or people who are likely to commit crimes. While some group members noted that this sort of prediction is beneficial when it focuses on the exploitation of minors, it can be more complicated when it is used in relation to sex workers. Although predictive data models may benefit some groups and individuals, possibly allowing authorities to prevent crime or to track and stop criminals, there are some crimes that are less black and white. Prediction can also apply on an individual level and may help individuals with self-management. If people would benefit from having specific types of information, it would be unethical to withhold it from them. However, are there cases where people may use this kind of data to draw inferences that would negatively affect them? Rather than making predictions using population-level data, it may also be about nudging individuals in a certain direction by identifying behaviors that may enhance health, are health neutral, or that tip towards health negative. Some workshop participants noted that recognizing these tipping points is crucial so that individuals are helped rather than harmed by predictive analytics and subsequent intervention.

When it comes to predictive analytics, the expected benefit is not always obvious and may differ depending on individuals' perspectives. Who is the person deciding what is beneficial? Some participants noted that it remains an open question whether some forms of predictive policing should be legal in the first place. One study demonstrates an ability to infer an individual's level of depression based on his or her social network activity. In this case, prediction is used in order to intervene in mental health issues, potentially benefiting people who are depressed. However, this may also have implications for human liberty and autonomy. On the other hand, giving someone predictive information about the onset of the flu may be helpful because she will be able to quarantine herself instead of infecting others.

Prediction implies certainty, but this is not always the case. While data may point to a particular child having some significant statistical likelihood of becoming a criminal, acting on this information this may also lead to false positives and other serious social or real harms. Predictive data models provide pieces of information, but individuals still need to decide what to do with this information. Someone still has to make a decision. How do you police the actions that you can predict when the predictions themselves are uncertain? Workshop members discussed various ways that modeling might be made more effective. Some of these possibilities include data collection standards that ensure data quality and anonymized datasets. Some workshop participants suggested that better modeling might prevent poor decisions from being made.

Transparency

Workshop participants discussed transparency at length, with some workshop members debating the merits and potential downfalls of transparency depending on context and use. First, it is imperative to note what transparency refers to. Are we talking about the data, the actors, or the information flow? Is transparency important for purposes of accountability? If we don't know what data we are talking about, who is using it, and to what end, it is impossible to make judgment calls about the possible benefit or risk of prediction. What if users were able to view their own data and also aggregate it in the same way as engineer might? Transparency may mean that individuals can see what is being done, i.e. how data is being collected and what sort of information is being tracked, as well as how the data will be used. It may also mean that individuals have access to how other people are being situated. Is it useful for people to be able to compare themselves to others, seeing how what the algorithms show them is different from what other people are seeing? When it comes to pricing, commercial entities do not have an incentive to offer transparency and consumers don't have easy ways to compare prices. What if algorithms select price differences based on factors like race? In one framing, this is an act of personalization but it might just as easily be called discrimination.

Multidimensionality, or the existence of multiple data sources over time, might undermine transparency. It is difficult to understand how specific browsing histories map to particular outcomes when there are so many factors at play. Data from multiple sources over time might impact the particular ad that a person sees. Can users see how all of these different factors and forms of data shape their experience? The definition of transparency is not just notice, but it also involves users' access to their own data. What is transparency intended for and who is granted access to this information? Transparency may in fact be beneficial, but it is not the solution in and of itself. It is possible, however, that it will create necessary pre-conditions that prevent discrimination from taking place.

Discriminatory algorithms

Workshop members debated the ways in which predictive data analytics can be used to discriminate. What happens when algorithms' predictive capabilities are used to punish or further disadvantage particular kinds of groups and individuals? If discrimination occurs, it may be difficult to implement a system that punishes a specific bad actor. It is tempting to focus on the human who is doing something wrong, but this kind of approach occludes systemic issues and may allow some individuals to game the system. When it comes to preventing discrimination through algorithm, there may be a need for targeted laws in certain sectors. Regulation targets specific cases that might have significant effects. We look at algorithms because we want to blame specific people, but should regulation try to target systemic effects rather than individual actors?

Group members addressed the problem of discrimination on individual and collective levels. Where is the line between the individual and the collective? What if group profiles are used to draw inferences about specific individuals? It is possible that researchers will lose a lot of information when they abstract up a level from the individual to a group, resulting in data that is inaccurate and overly clean. It is also possible that practitioners may shy away from aggregate data because they cannot make assertions about causation. Who should be making these inferences and does it matter if it is a person rather than an algorithm?

Discrimination itself is difficult to define. Modelers and legal entities may define discrimination differently. What counts as discrimination? Is it about funneling and personalization, in that certain individuals may have their choices limited? How do we update the existing law to deal with new forms of discrimination? For example, should we require the FDA and FTC to approve algorithms? There may be actors who fall outside of these types of regulation and are able to evade prosecution, so what should be done about these evasive entities?

Participants also discussed the problem of re-identification as a form of discrimination. Health data may benefit researchers and individuals, but there is also a risk of re-identification. Some participants argued that privacy is not the central concern when it comes to predictive data, questioning the significance of re-identification studies. How much is the interest in simply demonstrating the possibility of re-identification? If re-identification is an issue, does this mean that the solution is to abstain from publishing any data at all? What does it mean if individuals are re-identified in relation to sensitive health data and then potentially discriminated against?

Given that different individuals have different learning capacities to understand their own data, notice and consent agreements may not be sufficient. How can individual patients be contacted, informed, and agree to give consent before their data is used for predictive purposes? If individuals are told the ways in which their information might be used to form predictive models, what exactly should they be told and in what kind of detail? If we foist responsibility on the end user, it is possible that historically marginalized communities will suffer the most because they may lack the necessarily technical literacy to even recognize certain dangers. Such groups may be underexposed to certain technologies and may also lack the opportunities and resources necessary to familiarize themselves with them. This fact complicates the notion of consent and also outlines a new possible form of discrimination.

Further Exploration

Participants discussed how the legal, ethical, and social frameworks that were developed in the 1970s may no longer be applicable when it comes to present-day predictive data. How do we update these frameworks to maintain relevance in the face of immense technological changes? Not just the possibility for corruption or privacy risk, but the potential to create common good has expanded, too. How do we use these new technologies to maximize the common good?

Some members of the group discussed and flagged three main issues as subjects in need of further exploration. One such issue is framing. While some organizations, institutions, and individuals may consider prediction to be a matter of equity, others might view it as a justice or privacy issue. The ways that certain issues are framed, i.e. as matters of privacy, equity, or justice, will affect what appears to be at stake and the kinds of solutions that are deemed most appropriate. Privacy and equity are often conflated in these different framings. This may convolute the separate issues, but it is also maybe productive to think about how these different factors comingle. When we talk about the predictive power of data analytics, our conversations may address problems of manipulation, discrimination, or stigmatization. When particular aspects of these questions are emphasized, it may affect where remedial solutions are applied. What avenues are worth pursuing? Is justice just what someone deserves? While users may sometimes deserve the ability to make choices, sometimes they may also deserve to be protected from harm. These different scenarios need to be parsed.

Another issue that some workshop participants debated was the question of allocating burdens. What is transparency for? For whom is it for, and to what end? Some members of the group discussed the question of individual choice. How heavy of a burden can we place on individuals? Should the ones who act on data bear more of the burden? Is it the algorithm who is to be blamed for acts of discrimination or individual human actors? Even if you technically follow the rules, do you have additional responsibilities, obligations, or duties? What sorts of actors are actually covered by regulations? In selecting particular actors and parts of the ecosystem to be targets of regulation, that itself introduces questions of equality and fairness. There are cases of abuse or misuse, but there are also unintended consequences to prediction models that at first seem neutral or beneficial. Likewise, some intended outcomes may in fact be objectionable or raise new ethical questions.

The third issue that remains an open challenge is the question of existing infrastructures and actors that may address these problems. Are there existing infrastructures and regulatory agencies that might appropriately respond to the aforementioned challenges? Are there laws and agencies that can address these problems? What about civil society, community groups, and civil rights institutions? What does it mean to be a data caretaker in the context of existing social infrastructure?