Selected Papers of #AoIR2017:
The 18th Annual Conference of the
Association of Internet Researchers
Tartu, Estonia / 18-21 October 2017

Human Agency on Algorithmic Systems

Ansgar Koene
University of Nottingham

Elvira Perez Vallejos
University of Nottingham

Helena Webb
University of Oxford

Menisha Patel
University of Oxford

Abstract

The majority of large web service platforms promote service personalization as a means of providing user convenience. Parameter settings are commonly inferred from user behavior rather than explicit requests. As a consequence, users have little direct control of the system behavior. In this paper we discuss the implicit trade-off between user convenience and agency over the online experience. We will present results from multi-stakeholder engagement workshops exploring concerns and routes to solutions as raised by participants from industry, civil-society, teachers and academics. Finally, we will place these results within the context of the wider debate about User Trust on the internet and specific efforts to develop a standard on algorithm bias considerations.

Introduction/Context

From search engines to social-media feed personalization and news recommender systems, online access to information is heavily mediated by algorithmic systems that filter and guide the behavior of users. In the competition between services, the limited capacity of human attention is perceived as the main resource bottleneck and thus the primary factor in the competition between services. In response to this, increasingly

Koene, A., Perez, E., Webb, H., Patel, M. (2017, October 18-21). *Human Agency on Algorithmic Systems*. Paper presented at AoIR 2017: The 18th Annual Conference of the Association of Internet Researchers. Tartu, Estonia: AoIR. Retrieved from http://spir.aoir.org.

sophisticated and personalized algorithms are used to cut through the mountains of available information in the hope of providing content that is sufficiently enticing to keep the users on the platform. Superficially, there seems nothing wrong with prioritizing information that users will likely agree with; after all, people tend to self-select information that aligns with their own beliefs anyway. However, the implementation, and sometimes the very existence, of these personalization algorithms is often hidden from users with potentially negative consequences for their personal agency over their internet experience. Rather than ask users to explicitly define the topics or content the algorithm should select for, data mining and behavior tracking are used to algorithmically infer/identify personalized interest patterns. In order to perform this inference of personal interests the algorithms have to rely on certain basic assumption about user behavior, such as an assumption that browsing behavior is rationally efficient (e.g. time spent on a website is assumed to correlate with level of interest). Despite such assumptions the ‘big data’ properties of high volume and high dimensionality are assumed to produce better predictions of the user’s interests than even the users themselves are capable of consciously expressing [1].

When challenged over concerns that these highly complex algorithms are imposing editorial control over the information that people can access (e.g. amplifying the ‘echo chamber’ phenomenon), and should therefore require the service providers to be classified as media organizations with corresponding editorial responsibilities and regulations, the ‘data driven’ aspect of the algorithms is often used to argue that regardless of the filtering, ranking and recommendations performed by the algorithms, the platforms remain ‘content neutral’ technology providers, and are thus not media companies [2]. In brief, it is argued, that since the algorithm makes its inferences based on the data which is provided by the user it is not the service provider but rather the user who ultimately determines the behavior of the algorithm. Such an argument however is contestable when one considers that the users are often unaware of the types of actions that are monitored to provide input data for the algorithm, and have no way of knowing how those actions affect the algorithm behavior. Instead, the combination of design decisions by the service provider acting on behavioral data mined from the user results in a system that is not fully controlled by either party, with the potential to produce unexpected and undesired results. Many examples of unintended bias by information mediating algorithms against racial, gender or other protected groups are most likely the results of this kind of semi-blind interaction of applying a ‘black box’ system to an uncontrolled data set. When the Google Advertising algorithm showed more ads related to criminal background checking for typically African-American names than ‘white’ names [3], or showed higher paying jobs more to men than women [4], a general purpose ‘black box’ algorithm was probably mining historical data sets, that had not been screened against such bias, resulting in the perpetuation of unjustified bias.

Methodology

Our project [5] is working to develop tools, educational material and recommendations for regulatory safeguards to re-assert users’ agency over their online information access, with special focus on the identification and avoidance of unjustified algorithmic bias. The project combines deliberative discourse with 13-17 years old ‘digital natives’ with user behavior observation studies to map user experiences with algorithmic online systems, as well as experiments with algorithm design principles. The results of these studies are used as basis for engagement with stakeholders from industry, civil-society, government regulators and academia to develop recommendations for better education, design and regulation of these systems. The stakeholder engagement takes the form of workshop based discussions around specific case studies, e.g. the manipulation of page rankings in search engine results, and questionnaires. The discussions are recorded, transcribed and anonymized. The transcripts and the anonymized questionnaire responses are subsequently analyzed to identify common themes expressed by the participants.

Results

Outcomes of the first multi-stakeholder engagement workshop suggest that participants ranging from SMEs to NGOs and academics (legal and/or technical experts) consider a number of ‘agency’ related factors to play an important role in establishing the perceived ‘fairness’ of an algorithm. Factors such as: ability to appeal or negate an algorithm’s decision; freedom to explore algorithm effects by experimenting with the algorithm in a non-binding way; ability to have an explanation for the algorithm outcomes; ability to control the data that is used by the system; and an ability to explicitly ‘manually’ fine-tune the algorithmic system to match personal user objectives in recognition of the subjective nature of fairness. The next step in our exploration of user agency on algorithmic systems will consider the feasibility and desirability of implementing the proposed capabilities that were identified as ‘fairness’ enhancing factors.

Implications

On the policy/regulatory side, we are taking the results of our user and stakeholder studies to contribute to initiatives such as the User Trust agenda of the Internet Society [6] and the IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems [7]. As part of the latter initiative we are leading an IEEE Working Group for the development of a Standard on Algorithmic Bias Considerations [8]. This Standards is focused on ‘surfacing’ and evaluating societal implications of the outcomes of algorithmic systems, with the aim of countering non-operationally justified results. It will provide individuals or organizations creating algorithms with methods to provide clearly articulated accountability and clarity around how algorithms are targeting, assessing and influencing the users and stakeholders affected by the algorithm.

Acknowledgement

This work forms part of the UnBias project supported by EPSRC grant EP/N02785X/1. UnBias is an interdisciplinary research project led by the University of Nottingham in collaboration with the Universities of Oxford and Edinburgh. For more information about the UnBias project, see http://unbias.wp.horizon.ac.uk/

References

[1] Youyou, W., Kosinski, M. and Stillwell, D. (2015). Computer-based personality judgments are more accurate than those more by humans. PNAS, 112(4), 1036-1040.

[2] Segreti, G. (2016). Facebook CEO says group will not become a media company. Reuters, 29 August – http://www.reuters.com/article/us-facebook-zuckerberg-idUSKCN1141WN
[3] Sweeney, L. (2013). Discrimination in online ad delivery. Queue, 11(3), 1-19.
[4] Vincent, J. (2015). Google's algorithms advertise higher paying jobs to more men than women. TheVerge, July 7 – http://www.theverge.com/2015/7/7/8905037/google-ad-discrimination-adfisher
[5] UnBias http://unbias.wp.horizon.ac.uk/
[6] Internet Society (2016). A policy framework for an open and trusted Internet. White Papers, 22 June – https://www.internetsociety.org/doc/policy-framework-open-and-trusted-internet
[7] https://standards.ieee.org/develop/project/7003.html