Ethics of personalized information filtering

Ansgar Koene, Elvira Perez Vallejos, Christopher J. Carter, Ramona Statache, Svenja Adolphs, Claire O'Malley, Tom Rodden and Derek McAuley

HORIZON Digital Economy Research
University of Nottingham, UK
Email: first_name.last_name@nottingham.ac.uk

Keywords: Privacy, Transparency, Behavior manipulation, RRI, Filter bubble

Abstract. Online search engines, social media, news sites and retailers are all investing heavily in the development of ever more refined information filtering to optimally tune their services to the specific demands of their individual users and customers. In this position paper we examine the privacy consequences of user profile models that are used to achieve this information personalization, the lack of transparency concerning the filtering choices and the ways in which personalized services impact the user experience. Based on these considerations we argue that the Internet research community has a responsibility to increase its efforts to investigate the means and consequences of personalized information filtering.

1 Introduction

In the world of on-line information services the dominant business model is one in which no monetary payment is taken from the users. In order to attract maximum user numbers, information businesses therefore find themselves competing primarily based on the perceived quality of their information provision. Since information quantity is usually virtually limitless (which is why quantity is not a viable option for differentiating from competitors), information overload has become one of the main concerns for users. Perceived quality is therefore primarily determined by the ease with which the user can obtain some information that satisfies their current desires. The development of personalized information filtering therefore represents a logical step in the evolution of on-line information services. For many of the most highly success internet service, like Google, Amazon.com, YouTube, Netflix and TripAdvisor, the recommender system is a key element in their success over rival services in the same sector. Some, like Netflix, openly acknowledge this even to the extent of awarding large prizes for anyone that can improve their recommender system.

The simple logic behind the business case for developing such filtering systems however is not sufficient to put to rest the numerous social and ethical concerns that are introduced by the use of these filters. From a Responsible Research and Innovation (RRI) perspective [1], it is necessary for the internet research community to consider the wider implications of such innovations on society.

adfa, p. 1, 2011.

One of the social concerns about personalized information filtering that has probably attracted the most attention is the fear that optimizing people's information flows to focus on those things they have previously shown an interests/affinity for may cause a feedback loop by which people become isolated from new information due to a self-reinforcing filter bubble [2, 3]. To what extent this can, or does, happen as a consequence of search engine, social media and news feed personalized filtering, is not yet clear. While [4] provided a theoretical analysis showing that, under certain conditions, such a scenario is possible, little experimental work has been done to verify if the 'filter bubble' scenario is taking place. Under some circumstances, it was shown that a personalized Recommender System for music purchases appeared to widen the user's interests [5] rather than narrowing it. In this context it should also be noted that some recommender systems are being specifically designed to promote 'serendipitous discovery' [6].

Unfortunately, most of the research on the impact of personalization and recommender systems has so far focused on their commercial success in increasing sales (e.g. [7]), web impressions (e.g. [8]), and their ability to increase the consumer interest for niche goods (e.g. [9]). As we have argued in our previous position paper [10], this apparent imbalance in research efforts, seemingly focused on a corporate agenda, is exactly the kind of narrative that led to the GMO crop controversy in the EU in the 1990s which dramatically impacted the funding and public support for the Biosciences. In order to avoid such a public backlash against Internet research it is necessary to show that the research community is not solely interested in furthering a corporate agenda, but rather is seriously engaged with identifying and improving the societal impact of Internet research and innovation.

In the remainder of this paper we will focus on a number of other concerns associated with personalized filtering. The main social and ethical concerns we want to draw attention to in this paper are:

1. the privacy intrusion that is unavoidably linked to the tuning of the user behavior profile models;
2. the lack of transparency concerning the data that is used, how it is gathered and the way the algorithms work;
3. the risks of covert manipulation of user behavior.

2 Brief review of recommender systems

Recommender systems emerged as an independent research area in the mid-1990s. These first recommender systems [11] applied collaborative-filtering which works on the principle that a user who has in the past agreed with certain other users (i.e. given similar ratings, or 'clicked' on similar items) will have similar interests to them and will therefore find relevant and recommendations for items that these users rated highly. Modern recommender systems using (combinations of) various types of knowledge and data about users, the available items, and previous transactions stored in customized databases. The knowledge and data about the users is collected either through explicitly ratings by the users for products, or are inferred by interpreting user

actions, such as the navigation to a particular product page which is interpreted as an implicit sign of preference for the items shown on that page.

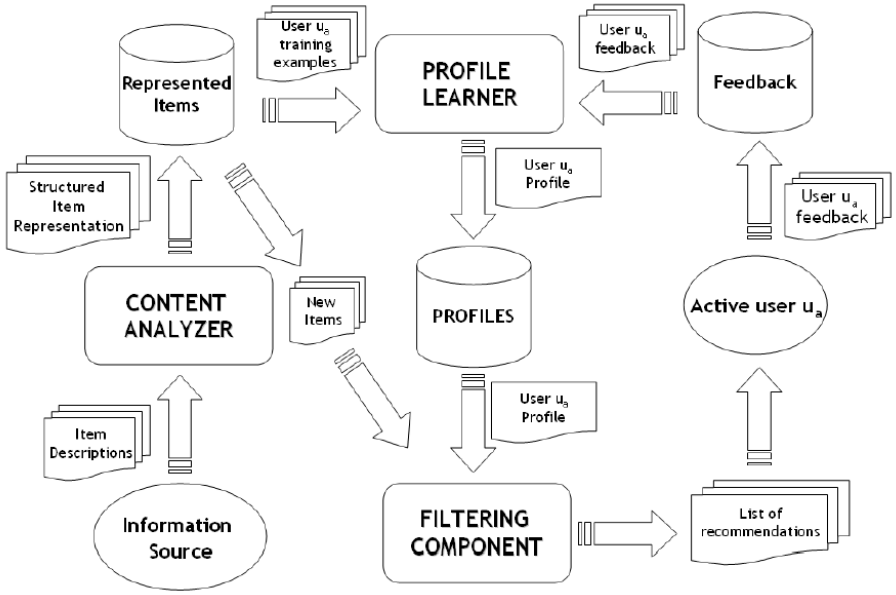
The two main classes of recommender systems are:

- Content-based, where the system learns to recommend items that are similar to the ones that the user liked in the past. The similarity of items is calculated based on the features associated with the compared items. Figure 1 gives a high-level overview of the components and data flow in a content based recommender system.
- Collaborative-filtering, users are given recommendations for items that other users with similar tastes liked in the past. The similarity in taste of two users is calculated based on the similarity in the rating history of the users.
- Community-based, where the system recommends items based on the preference the user's friends. This is similar to Collaborative filtering except that the selection of peers to be used for identifying the recommendation is based on an explicit 'friendship' link instead of being deduced from patterns of similar past behavior. Such 'social recommender' systems are poplar in social-network sites [12].

In practice many of the recommender systems are hybrid systems that try to balance the advantages and disadvantages of each class [13]. Collaborative and community based systems, for instance, suffer from an inability to recommend items that have not yet been rated by any of the potential peers of the user. This limitation however does not affect content-based system as long as the new item is supplied with a description of its features, allowing it to be compared to other items that the user has interacted with in the past.

A comprehensive introduction to recommender systems is provided in [14].

Fig. 1. High level architecture of a Content-based recommender system



3 Privacy intrusion

The filter parameters that determine the personalized selection and ranking of information constitute an implicit user profile model, which is usually based for a large part on data about the past search and browsing behavior of the users when they previously interacted with the service [15]. To further refine the user profiles, services may also gather information about the user behavior on other websites through the use of 'tracking cookies' [16] or by purchasing third-party access to such data from other services. Additionally, recommender systems may also use data concerning the behavior of people within the social network of the users [17].

From a privacy and digital human rights perspective, each of these data gathering methods is ethically troubling since they are all surreptitious, to varying degrees. The use of 'tracking cookies' is clearly the most troubling in this respect, however even the logging of users' behavior when they are actively engaging with the information service itself lacks proper informed consent. At best, users may have read something about data logging in the terms-and-conditions they had to agree to when they first signed up to the service. Unfortunately the current reality of Internet usage is that terms-and-conditions policies of Internet sites are rarely read and are generally formulated in ways that are too vague and incomprehensible to constitute a means of gaining true informed consent [18]. Furthermore, it is not realistic to expect users to remain vigilantly aware of the information about data tracking in the terms-and-conditions weeks, months and years after they signed up to the site. The "EU Cookie Law" [19] has gone some way towards providing a frequent reminder of data tracking by websites, however the standard notification of the type:

"By continuing to use this site you consent to the use of cookies on your device as described in our cookie policy unless you have disabled them. You can change your cookie settings at any time but parts of our site will not function correctly without them." [FT.com]

is generally too vague for people to understand, resulting in the same dismissive 'click to close' behavior that people have become accustomed to from the cryptic error/warning messages that are typically generated by software and the terms-and-conditions agreements they did not read.

Beyond the data collection process, the user profile models that form the basis of personalized information filtering pose an additional privacy concern in themselves. The profile models of users are in essence an operationalization of the data mining efforts by the service provider, built to anticipate the user's behavior, interests and desires. Access to a perfect behavior model of a user would in principle enable anyone to predict the user's actions/decisions for a wide range of choices/conditions. Beyond the immediate commercial potential for guiding 'relevant' advertisements to a person, such person profiles could be used to plan targeted phishing campaigns or hacking related social engineering. To find a user's weaknesses it would suffice to query the user's behavior model with a range of choices and observe the predicted responses.

Based on the analysis above, we observe that models of user behavior profiles are necessary for the functioning of personalized recommender systems and that these

models are unavoidably linked to a certain level of privacy intrusion. We therefore propose that, within the RRI framework, the Internet research community should focus not only on ways to further fine-tune the recommendations from such systems, but also on developing recommender systems architectures where the user profile model, and the corresponding privacy sensitive information, stays within the access-control domain of the user. One possible method for this might be based on a two-layer architecture where the first layer, hosted by the information service provider, generates non-personalized search/recommendation results, which are provided to a second layer, hosted locally on the user's device, which ranks the results based on the personalized user profile.

4 Transparency

Due to the commercial advantage which information service providers hope to achieve through the use of personalized information filtering, the information about how exactly their filtering is done is not made publicly available. This lack of transparency however also makes it impossible for users to gain a full understanding of how their data is being gathered and used, thus preventing them from truly providing informed consent. A key concern in this regard is the fact that most of the information service businesses do not earn their money from the users but rather from corporate customers who pay access to user data in order to push targeted advertising. There is thus a valid argument for demanding greater oversight into functioning of the information filtering in order to guarantee that the information provided to users best servers their needs. Due to the personal nature of the data used for the behavior profiles, and the user models themselves, there also needs to be transparent verification that these are stored and handled in accordance with the jurisdictionally appropriate privacy-related regulation and laws (e.g. [20][21]). While this might be viewed as primarily the responsibility of regulatory oversight authorities, we propose that this should also be seen as a challenge to the Internet research community to develop tools with which the filtering criteria can be probed without access to the underlying code/algorithm. An obvious approach for developing such tools might be to follow existing Black-box testing practices which are commonly used in software and general systems development for evaluation and functional testing [22]. An example for this is provided by [23], where a Black-box testing approach was used to investigate what kind of recommendation schemes were exploited by various movie recommendation systems.

In accordance with the principle of public engagement and dialog concerning research and innovation, the RRI framework also suggests that the Internet research community should make user-friendly versions of recommender system testing kits available to the general public to enable people to evaluate for themselves if they find the level of profile personalization used by the recommender system acceptable of not.

5 Behavior manipulation

Directly relating to the concerns about lack of transparency, as well as the issue of 'filter bubbles' is the question of how much and for which agenda user behavior is being manipulated by the use of personalized filtering. To a certain extent, behavior manipulation is unavoidable in any information presentation system since people will invariably select the first items on a list more often than those much further down. Since it is impossible to place all information at the top of the list, the act of ranking involves a behavior manipulation. Provided the manipulation is based on mutual consent, there is nothing wrong with this. It is in fact the desired function of a search or recommendation system, as long as the user knows and agrees to the ranking criteria that are used by the algorithm. Advertising, of course is all about attempting to persuade, i.e. manipulate, potential consumers into purchasing the product/service of the advertising agency's client. The dominant business model of advertising funded online information service therefore constitutes a significant conflict of interests at the heart of the filter criterion selection.

As long as the information filtering, and advertising targeting, were based on global statistical criteria it was usually relatively easy for users to judge if the information they were provided with was advertising motivated. The general coarseness of the match between the provided and the desired information also meant users engaged a more critical attitude towards evaluating the search results. The introduction of personalized information filtering however is improving the personalized targeting success of advertisements at least as fast as the general filtering success. Also, due to the generally improved information services, people are less critical in their final selection.

In case there was any doubt, the willingness of information service providers to engage in manipulating their information filtering for purposes other than the service to the user was clearly demonstrated by the "Facebook news feed experiment" [24].

Once again, there is undoubtedly a role for regulatory oversight concerning these conflicts of interest, similar to such regulation in other media. The fact that the personalized filtering and advert targeting systems are developed by Internet researchers, however, means that the Internet research community will undoubtedly be implicated in any future scandals about manipulation of personalized information filtering, as it already was with [24]. In order to mitigate the impact of such events it is therefore important for the Internet research community to be visibly engaged with RRI agenda.

The challenges in this case are simultaneously daunting and yet very familiar: how to prove beyond reasonable doubt that a statement/recommendation is objectively true and unbiased. Following the example of legal court cases, the first step might be to ask the defendant, i.e. the service provider who controls the recommender system, to provide evidence concerning the basis for the recommendation.

In a bid to gain people's trust, and boost interest in the offered recommendations, various recommender systems, e.g. Amazon.com, already provide some level of evidence by informing the users why certain recommendations are given with statements like: "Customers Who Bought This Item Also Bought". While such information can clearly help users to better understand, and thus evaluate and trust, recommendations

it does not fully address concerns about possible behavior manipulation. Such concerns can only be addressed through the provision of trusted-third-party involvement, either in the form of regulatory oversight or by providing tools which can allow users to test the recommender system for un-accounted for recommendation biases.

6 Evidence of public concern about recommender systems

In this section we summaries a number of news stories that illustrate the level of concern, rightly or wrongly, over lack of transparency and potential bias in recommender systems.

In 2011, the US Federal Trade Commission started an investigation into possible search results bias by Google. It took two years of investigating before Google was cleared of the charges [25].

In February 2014, Google agreed to a settlement with European competition regulators following years of legal struggles with antitrust authorities, starting in 2010, concerning complaints that Google search rankings unfairly favored Google products [26].

In 2010, Netflix decided to cancel the Netflix Prize sequel after the US Federal trade Commission raised concerns about Netflix user privacy and a lawsuit was filed against Netflix [27]. The Netflix Prize competition, and its planned sequel, challenged competitors to develop improved recommendation algorithms based on a published set of anonymized Netflix user data of the type. One of the reasons for the privacy concerns was the publications in 2008 of a paper showing that the data supplied for the recommender algorithms by the Netflix prize dataset was rich enough to allow it to be de-anonymized [28].

7 Conclusion

Personalized information filtering by online search engines, social media, news sites and retailers represents a natural evolution in the development towards ever more finely tuned interaction with the users. Even leaving aside concerns about individual and social consequences of possible 'filter bubbles', the user profiling required to achieve this personalization raises numerous ethical issues around privacy and data protection. Further concerns arise due to the lack of transparency and the potential for increasingly covert manipulation of user behaviour in favour of the commercial interests of the predominantly advertising based business models of information services. Due to the frequently close involvement of the large information service providers with the Internet research community, there is a growing risk that scandals related to personalized information filtering by corporations might triggering a controversy and public backlash similar to the one that hit GM crops in Europe in the 1990s. In order to avoid such a controversy it is essential to retain the confidence and trust of the public by actively engaging with the Responsible Research and Innovation agenda and pro-actively working to mitigate these issues. In order to achieve this we propose a research programme aimed at:

- identifying and studying the socio-psychological impact of personalized filtering;
- helping people to understand and regulate the level of privacy intrusion they are willing to accept for personalized information filtering;
- developing a methodology to probe the subjective 'validity' of the information that is provided to users based on their own interests;
- engaging with corporate information service providers to reinforce ethical practices.

Project elements for such a research programme might include

- Technical development of tools:
  - Black-box testing kit for probing the characteristics of the user behavior profiles used in recommender systems.
  - Recommendation bias detection system for identifying user behavior manipulation
  - A two-layer recommender architecture that de-couples the delivery of non-personalized information by service providers from a user owned/controlled system for personalized ranking of the information.

- Psycho-social research on the impact of personalized information filtering on:
  - General exploration-exploitation trade-off in action selection
  - Attitudes towards trust and critical evaluation of information

- Cybersecurity:
  - Protection against mal-use of personalized recommender systems for phishing related social engineering

- Policy:
  - Development of guidelines for responsible innovation and use of recommender systems, protecting the privacy and freedom of access to information of users.

- Public engagement:
  - Develop educational material to help people understand how recommendations they receive from search engines, and other recommender systems, are filtered so that they can better evaluate the information they receive.

8 Acknowledgement

This work forms part of the CaSMa project supported by ESRC grant ES/M00161X/1. For more information about the CaSMa project, see http://casma.wp.horizon.ac.uk/ .

9 References

1. H. Sutcliffe, "A report on Responsible Research & Innovation" in *Matter*, 2011. Obtained through the internet http://ec.europa.eu.
2. E. Pariser, *The Filter Bubble*, Penguin Books, 2011.
3. C. R. Sunstein, *Republic.com*, Princeton: Princeton University Press, 2007.
4. M. Van Alstyne and E. Brynjolfsson, "Global village or cyber-Balkans? Modeling and measuring the integration of electronic communities", *Management Science*, 51(6):851-868, 2005.
5. K. Hosanagar, D. Fleder, D. Lee and A. Buja, "Will the Global Village Fracture into Tribes? Recommender Systems and their Effects on Consumer Fragmentation", *Management Science*, 60, 805-823, 2014.
6. Y. Cao Zhang, D. Ó Séaghdha, D. Quercia, T. Jambor, "Auralist: Introducing Serendipity into Music Recommendation", *Proc. of the 5th ACM Int. Conf. on Web Search and Data Mining (WSDM'12),* February 8–12, 2012, Seattle, Washington, USA
7. P. De, Y. J. Hu and M. S. Rahman, "Technology usage and online sales: an empirical study", *Management Science*, 56(11):1930-1945, 2010.
8. A. Das, M. Datar, A. Garg, and S. Rajarm, "Google news personalization: scalable online collaborative filtering", *Proc. of the 16th Int'l World Wide Web Conference*, 271-280, 2007.
9. D. Fleder and K. Hosanagar, "Blockbuster culture's next rise or fall: the impact of recommender systems on sales diversity", *Management Science*, 55(5):697-712, 2009.
10. A. Koene, E. Perez, C. J. Carter, R. Statache, S. Adolphs, C. O'Malley, T. Rodden and D. McAuley, "Research Ethics and Public Trust, Preconditions for Continued Growth of Internet Mediated Research", at *1st International Conference on Information System Security and Privacy (ICISSP)*, 9-11 February, 2015.
11. D. Goldberg, D. Nichols, B.M. Oki, D. Terry, "Using collaborative filtering to weave information tapestry", *Commun. ACM*, 35(12), 61–70, 1992.
12. J. Golbeck, "Generating predictive movie recommendations from trust in social networks", *Trust Management, Proceedings 4th International Conference, iTrust 2006*, Pisa, Italy, 93–104, May 16-19, 2006.
13. R. Burke, "Hybrid web recommender systems", *The AdaptiveWeb*, 377–408. Springer Berlin / Heidelberg, 2007.
14. L. Rokach, B. Shapira, and P.B. Kantor. *Recommender systems handbook*. Vol. 1. New York: Springer, 2011.
15. M. Speretta, S. Gauch, "Personalized search based on user search histories," *Web Intelligence, 2005. Proceedings. The 2005 IEEE/WIC/ACM International Conference on*, vol., no., 622-628, 19-22 Sept. 2005. doi: 10.1109/WI.2005.114
16. T. Rohle "Desperately seeking the consumer: Personalized search engines and the commercial exploitation of user data". *First Monday***,** [S.l.], sept 2007. ISSN 13960466. <http://journals.uic.edu/ojs/index.php/fm/article/view/2008/1883>
17. H. Ma, D. Zhou, C. Liu, M. R. Lyu and I. King, "Recommender systems with social regularization", *WSDM '11 Proceedings of the fourth ACM international conference on Web search and data mining*, 287-296. 2011. doi: 10.1145/1935826.1935877
18. E. Luger, "Consent for all: Revealing the hidden complexity of terms and conditions", *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2687-2696, 2013.
19. The Privacy and Electronic Communications (EC Directive) Regulations 2003, available at http://www.legislation.gov.uk/uksi/2003/2426/contents/made

20. European Union (EU) Data Protection Directive of 1995 (Directive 95/46/EC), available at http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:31995L0046
21. OECD Guidelines on the Protection of Privacy and Transborder Flow of Personal Data (C(80)58/FINAL, as amended on 11 July 2013 by C(2013)79), available at http://www.oecd.org/sti/ieconomy/privacy.htm
22. B. Beizer, *Black-box testing: techniques for functional testing of software and systems*. John Wiley & Sons, Inc., 1995.
23. N. Lee, J.J. Jung, A. Selamat, and D. Hwang, "Black-box testing of practical movie recommendation systems: A comparative study", *Computer Science and Information Systems, 11*(1), 241-249, 2014.
24. A. D. I. Kramer, J. E. Guillory and J. T. Hancock, "Experimental evidence of massive-scale emotional contagion through social networks", *PNAS* 111, 24, 8788-8790, 2014.
25. C. Arthur, "Google cleared of search results bias after two-year US investigation", *the Guardian*, January 4th, 2013
26. C.C. Miller and M. Scott, "Google Settles Its European Antitrust case; Critics Remain", *the New York Times*, February 5th, 2014.
27. Netflix official blog announcement, March 12, 2010. http://blog.netflix.com/2010/03/this-is-neil-hunt-chief-product-officer.html
28. A. Narayanan and V. Shmatikov, "Robust de-anonymization of large datasets (how to break anonymity of the Netflix prize dataset)", *University of Texas at Austin*, 2008.