

Behaviorism is Not Enough: Better Recommendations through Listening to Users

Michael D. Ekstrand^{1,2} and Martijn C. Willemsen³

¹Dept. of Computer Science
Texas State University
San Marcos, TX USA

²Dept. of Computer Science
Boise State University
Boise, ID USA
michael.ekstrand@
boisestate.edu

³Human-Technology Interaction Group
School of Innovation Sciences
Eindhoven University of Technology
Eindhoven, The Netherlands
M.C.Willemsen@tue.nl

ABSTRACT

Behaviorism is the currently-dominant paradigm for building and evaluating recommender systems. Both the operation and the evaluation of recommender system applications are most often driven by analyzing the behavior of users. In this paper, we argue that listening to what users say — about the items and recommendations they like, the control they wish to exert on the output, and the ways in which they perceive the system — and not just observing what they do will enable important developments in the future of recommender systems. We provide both philosophical and pragmatic motivations for this idea, describe the various points in the recommendation and evaluation processes where explicit user input may be considered, and discuss benefits that may result from considered incorporation of user preferences at each of these points. In particular, we envision recommender applications that aim to support users' *better selves*: helping them live the life that they desire to lead. For example, recommender-assisted behavior change requires algorithms to predict not what users choose or do now, inferable from behavioral data, but what they should choose or do in the future to become healthier, fitter, more sustainable, or culturally aware. We hope that our work will spur useful discussion and many new ideas for recommenders that empower their users.

Keywords

Recommender systems; participatory design; user studies; algorithmic filtering; information filtering

1. INTRODUCTION

Recommender system developers rely on two different types of data in order to power and evaluate recommenders: expressed preferences, including data like user ratings, survey responses, and other explicit judgements from the user; and behavioral data or implicit feedback, arising from recommendation interactions and other actions the user takes on the system. Recommender systems have long used both types of data: early work typically used ratings for recommendation and a combination of user input and behavior observation to evaluate the system [9].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

RecSys '16, September 15 - 19, 2016, Boston, MA, USA
Copyright is held by the author(s). Publication rights licensed to ACM.
ACM 978-1-4503-4035-9/16/09...\$15.00
DOI: <http://dx.doi.org/10.1145/2959100.2959179>

In recent years, behaviorism has ascended to be the dominant paradigm for recommendation. This emphasis takes many forms, but is particularly visible in the trends of favoring implicit feedback over explicit ratings, behavioral evaluations such as A/B tests measuring user activity instead of surveys or ethnographic analyses that capture users' experience and perception of the system, and the practice of ignoring preference when it disagrees with behavior.

User behavior is a valuable source of input. Implicit feedback data is readily available — soliciting ratings or responses from users is a challenging task — and is very good for predicting future user behavior. Behavior-based evaluations directly measure users' visible responses, which translate to short-term revenue-generating actions such as purchases.

We argue that this is not enough. While behavioral data rightly plays a significant role in recommendation, we must not neglect users' stated preferences and ambitions. Further, we submit that eliciting stated desires from users has great potential to enable uniquely empowering recommendation experiences.

Over the last half-decade, significant progress has been made on developing evaluation strategies that integrate behavioral and subjective data. These methods are effective for taking advantage of the relative strengths of both classes of data and form a starting point for the types of evaluations and systems that we envision.

At present, recommender systems research has yielded an extensive knowledge base of how to produce effective recommendations that will prompt users to action. But there are several open questions that we must engage users, explicitly, to address:

- Are users satisfied with their choices and actions, both in short term as well as in the longer term?
- When a user is dissatisfied with their choices, what keeps them from aligning their actions and desires? Can the recommender suggest alternatives, or help the user explore the item space or express preferences in a different way?
- Is the recommender feeding users' better or lesser selves? Is it providing recommendations that help the user achieve the goals they have for their life, or is it reinforcing behavior that detracts from those goals?

In this paper, we first outline some philosophical and pragmatic motivations for increasing user participation in the recommendation process. We will then examine where in the recommendation process explicit user input could be used, and propose certain specific benefits we see for increased use of user input in each. We hope the arguments we advance will spur useful discussion about how to design, build, and study recommenders that enable users to find material and courses of action that empower them to live more fulfilled and fulfilling lives.

2. MOTIVATING USER PARTICIPATION

Our motivations broadly divide into two categories: a values-based philosophical argument for why user participation is an intrinsically good objective and a pragmatic argument centered around the kinds of applications that a new focus on user input may enable.

2.1 The Philosophy of User Participation

Users and system developers often have different visions. Even when the designer's vision is supported by empirical data on user *behavior*, the result can still be poorly received if the users feel like their values are not being respected, or that they have no control over the technology in their lives.

Franklin advocates for a recognition of the role of *reciprocity* in our technological development [7]. Reciprocity — the give-and-take present in egalitarian face-to-face human interaction, a 'genuine communication among interacting parties' — is important for technology to empower, rather than dominate, its users. Without reciprocity, users can only take or leave what is given to them.

A major theme of Franklin's work is the need to carefully consider the values embedded in our technologies. Suchman strikes a similar note when laying out the merits of participatory design, stating that it 'makes explicit the critical, and inevitable, presence of values in the system design process' [21].

Values are present in all system designs. The question isn't *whether* a system embeds some set of values; rather, *whose values* (and *what values*) does it embed? Are these values explicitly articulated and subject to discussion? Are designers transparent about the reasons for decisions? Giving credit to user perspectives is itself a value that can be included in or rejected by a technological process, and including user input can enable debate and discussion about the other values the system embodies.

The values of recommender systems are seen most clearly in our optimization criteria and evaluation metrics. Sales, lift, engagement, click-through rate: all of these are value judgements about the kinds of activities that reflect a good recommendation (and of whether a recommender should even be deployed). Does engagement, for example, truly reflect what users need from a service [6]?

The recent debate over Twitter's introduction of algorithmic filtering to user timelines highlights how conflicts over values can arise in recommender systems. Changing from the traditional reverse-chronological timeline to a recommender-filtered one was a significant change in core service functionality, and it displeased a number of vocal users. One of the common complaints was that Twitter did not seem to understand how existing users used its system and the actual problems they faced. If users felt heard, and like their desires and problems were respected, might Twitter's algorithmic filtering have been better received?

The issue of *filter bubbles*, brought to public attention by Pariser [15], is another matter of concern for some users of modern information systems: does algorithmic filtering isolate us into pockets of one-sided information, to the detriment of ourselves and society?

Suchman observes that 'until we become familiar with and take seriously each other's concerns there will be little hope for a mutually satisfactory future in the development of work and technology' [21]. Giving credit to user concerns is a key dynamic in building technology that truly satisfies those it impacts.

2.2 Enabling New Applications

Besides our philosophical concerns, there are several practical reasons why increasing the role of explicit user input of preferences and goals can enable significant advances in recommender systems.

First, behavioral data without proper grounding in theory and in subjective evaluation might just result in local optimization or short term quick wins, rather than long term satisfaction. Further, user behavior may be driven by outside factors, such as a lack of meaningful alternatives. When can we know from the behavior of a user if the recommendations help to fulfill their needs and goals?

Neil Hunt raised a striking example in his keynote at RecSys 2014. Netflix relies heavily on behavioral data for evaluating recommenders and other service components, but their key metric — users watching movies — cannot distinguish between a user deriving value from the service and addiction. The recommender, which helps many users find movies and shows that enrich their lives, may be doing some users a disservice by encouraging them to continue watching to the neglect of other important concerns in their lives. Engaging users about their concerns and goals can be a valuable tool in mitigating this kind of problem in recommender systems.

This concern is especially relevant in the emerging area of lifestyle modifications and behavioral change. Recommending the kinds of things users like or do now will not help them to achieve the behavior they want, but rather will reinforce their current behavior. These systems require more than implicit feedback, and optimization for user needs beyond immediate satisfaction. For example, the user model in such a system needs to understand the users' goals or desires and the algorithm needs to optimize for attaining them.

There are certainly many other applications that can be envisioned, but several of these ideas have a common theme: can the recommender help users become their better selves?

3. WHERE USERS CAN SPEAK

We now survey the different points in the recommendation process where explicit user input can be integrated, and propose applications that could benefit from increasing the role of explicit user input in these areas.

3.1 Application Design

Participatory design itself is most relevant in the process of designing a recommender application and its algorithm(s). In this phase, users can express what they want the recommender to help them accomplish. What do they desire from news recommendations? People recommendations? Algorithmic sorting of status updates? Recipes? Consumer products?

User input must be balanced with product vision, particularly in the initial design stage; we are not arguing against a need for expert judgement in design. Moreover, users might not know what they want (as they might not be aware of their possibly-latent needs), but they can respond to ideas, sharing what they like or dislike, and might be prompted to come up with even better ideas. We also do not oppose the use of behavioral data to influence design — the rise of A/B testing for design as opposed to relying entirely on potentially-unsubstantiated judgement is welcome. We are arguing, rather, for including users' expressed desires as a third leg in the design process of recommenders and related systems.

3.2 Recommendation Method Selection

Recommenders can be designed so users have control over which of several algorithms is providing them with recommendations. This is a direct means of giving users agency in the recommendation process, and with suitable explanations of the algorithms it can allow the users to adapt the system to their goals at any given time.

There has been some work on allowing users to choose their recommenders [5], but there is much more to be done in understanding how we can provide users with meaningful control over the way in which their recommendations are produced. We could envision

systems where users are provided with a number of recommendation techniques that they can mix and match in order to achieve a particular type of recommendation list; a more sophisticated version, perhaps, of the diversity control proposed by Ziegler [22] or the user adjustments provided by MovieLens [8].

Different recommenders have different characteristics, which can make them more or less effective for meeting various user information needs [4, 13, 14]. One approach to this varying effectiveness is to attempt to automatically adapt the recommender system to the user's need [10]. This can be thought of as a *recommender of recommenders*, where one algorithm 'recommends' the best recommendation algorithm to use in a particular context or task; this is one form of context-aware recommendation [2]. Another approach is to put the user in control of the process, allowing them to directly or indirectly determine the algorithm used. This has a number of benefits, including increasing user agency and collecting data that can be used to test and develop future meta-recommenders.

3.3 Recommendation Process

Once the recommender system has been designed and its algorithm selected, the ongoing process of recommendation depends on input data from the user. This typically comes in two forms: explicit ratings or preference judgements; and implicit feedback in the form of clicks, purchases, and other user events. While explicit ratings dominated the early days of recommender systems, much recent attention has focused on implicit feedback. The reasons for this are fairly clear: implicit data is far more plentiful, as it can be gleaned from users' ordinary actions instead of requiring them to take time to express preference, and the resulting recommendations produce better user response in many practical applications.

Drawing from experience at Netflix and elsewhere, Amatriain noted that implicit signals are better than explicit ones 'almost always' [1]. Indeed, many *current* applications (especially given their limited domains of multimedia or e-commerce) do fare better with implicit data. But we see great potential in exploring the cases outside of 'almost always': in what settings does explicit user input result in a better recommender system?

One example where it has been shown to result in better recommendations is Buzzr [16]. They studied several models for news item recommendation, and found that taking into account the user's explicit subscriptions produced more useful recommendations than recommending from a broader pool of feeds.

User input need not be complicated. Adding a 'must read' flag to Facebook or Twitter, so that users can indicate they want to read everything posted by certain friends, can help them retain control.

These examples show that explicit user input could improve some types of recommendations. But we believe that there are applications where explicit input beyond behavior is likely necessary, not merely beneficial, and it can be particularly useful to consider the *difference* between actual behavior and stated preference. When a user says one thing and does another — which they often do, as Hunt and others have observed — there are two possibilities:

- *The user does not understand their true desires.* In such cases ignoring their stated desires and inferring preference from their behavior will result in good recommendations.
- *The user is not satisfied with their behavior and may wish to change it.* In these cases, recommendations based on past user actions will only reinforce behavior the user wishes to change.

Applications where past behavior seems insufficient include:

- Finding speakers who bring perspectives that have been underrepresented in previous conferences.

- Reducing the stress and angst of a social media feed overpopulated with 'hate-reads'.
- Selecting high-impact next tasks from a to-do list.
- Changing musical or cinematic tastes.
- Eating healthier foods.
- Changing energy consumption behavior.

Some of these applications have a specific behavioral change component. To support such use cases, we believe it is useful to draw measures and concepts from social psychology. Behavioral change has been extensively studied by social psychologists, who have developed and tested social-cognitive models to describe how people can put good intentions into action.

To illustrate, many theories in this area assume that users have specific goals, and that implementation plans or actions are needed to help them overcome barriers to achieving these goals. For example, the Transtheoretical Model [17, 18] describes behavioral change as a 5-stage process: it starts with *precontemplation* and proceeds through *contemplation*, *preparation*, *action*, and finally to *maintenance*. The earlier stages require awareness, whereas the later stages require motivation and commitment. In each stage, a recommender might help with different suggestions: raising awareness may help the user in the precontemplation stage, whereas a personalized action plan might help support the preparation stage. Once the change is in maintenance mode, the recommender should avoid suggestions that might prompt the user to regress.

There are many research questions that arise from the goal of building such a system. How can we represent such action plans, stages and goals in a recommender system algorithm? What kinds of interfaces do we need to elicit the user's goals and present recommendations that will help them achieve them?

Providing adequate recommendations in this area has not yet been extensively studied. In the related area of Persuasive Technology a lot of progress has been made using the persuasion principles by Cialdini [3], but these attempts focus mostly on the how to persuade, not what to persuade with. Personalization is done on the message [11] but not yet on its content. However, effective e-coaching should not only address the how, but should also understand the what and where [19]. The 'what' (personalized recommendations) and on what moment (context awareness) are clearly challenges that recommender systems can take up.

For one example of how psychological theory can guide the structure of an algorithm, Starke et al. [20] designed an energy-saving recommender system based on a Rasch scale, which orders behaviors according to how costly or difficult they are to implement. For example, turning off the lights when leaving a room is easier than buying an energy-efficient washing machine. The Rasch model can also represent levels of ability on the same scale enabling the recommender to suggest energy-saving measures that fit a persons' ability. Moreover, the ordering of the scale can represent a hierarchy of goals the user might wish to attain in the future on their quest for their better self.

3.4 Evaluating the Recommender

Finally, we argue that explicit user input is useful in evaluating recommenders. A/B tests, the common gold standard for recommender assessment, are efficient and effective for testing the impact of different algorithms or designs on user behavior but are often limited in their ability to explain *why* users acted in a particular way. In addition to grounding experiments in theory, collecting subjective responses from users can help a great deal in explaining behavior.

User studies [12] are one means of doing this. Soliciting the user's subjective perceptions of the recommendations provides valuable

insight into their thought processes, and allows us to better understand how algorithmic concepts of constructs such as diversity do or do not map to users' perceptions of those concerns [4].

User studies are not a panacea for good research. Well-designed A/B tests can test theories and lead to generalizable knowledge, and poorly-conceived or underpowered user studies can produce little of lasting value; further, good studies are in general difficult to design, execute, and analyze, so the tool is not responsible for weak results. That said, it is *easier* to design surveys whose results provide theoretical insight into human behavior than it is to design strictly behavioral studies, because the link between psychological processes and observable outcomes is more direct.

In addition to survey-based user studies, qualitative and ethnographic studies can elicit how the users feel about the recommender system, its embedding application, and the way that interacts with the rest of their lives.

Thus, to advance our knowledge about *how* recommendations are received by their users, and the various psychological factors that affect their suitability, it is necessary to conduct many targeted user studies with explicit, subjective responses, not just behavioral studies, and to integrate these with qualitative analysis of user desires and choices. Further, we expect that deeper understanding of users' subjective experience of recommendation can improve commercial recommendation applications.

4. CONCLUSION

In this paper, we have laid out several ways in which explicit user input can be harnessed to improve recommender applications.

There's a great deal of research needed in order to enable future applications to fully realize this potential. Several important questions that arise from our argument include:

- How can we scale participatory design to Internet-scale applications?
- What does participatory design of algorithms look like?
- How can we harness the discrepancies between what users say and what they do to provide recommendations that might help bridge that gap, and aid the user in living the life that they wish to live?
- How can we provide meaningful control over the recommendation process to users, so that they can understand the decisions they make about their recommendations and customize the system to their particular needs?

Investing significant effort in harnessing the power of user input and explicit preference judgements has the potential to enable substantial new recommender systems that empower and enrich users' lives. Just as importantly, it can help us to maintain user agency in an increasingly algorithm-guided information world.

ACKNOWLEDGEMENTS

We thank Jennifer Ekstrand for her valuable insights and feedback on draft versions of this paper.

REFERENCES

- [1] Amatriain, X. 2016. Past, Present, and Future of Recommender Systems: An Industry Perspective.
- [2] Anand, S.S. and Mobasher, B. 2007. Contextual Recommendation. *From Web to Social Web: Discovering and Deploying User and Content Profiles*. B. Berendt, A. Hotho, D. Mladenic, and G. Semeraro, eds. Springer Berlin Heidelberg. 142–160.
- [3] Cialdini, R.B. 2001. *Influence: Science and Practice*. Allyn and Bacon.
- [4] Ekstrand, M.D., Harper, F.M., Willemsen, M.C. and Konstan, J.A. 2014. User Perception of Differences in Recommender Algorithms. *Proc. ACM RecSys '14*, 161–168.
- [5] Ekstrand, M.D., Kluver, D., Harper, F.M. and Konstan, J.A. 2015. Letting Users Choose Recommender Algorithms: An Experimental Study. *Proc. ACM RecSys '15*, 11–18.
- [6] Evans, J. 2014. Twitter's Huge Mistake. *TechCrunch*.
- [7] Franklin, U.M. 2004. *The Real World of Technology*. House of Anansi Press.
- [8] Harper, F.M., Xu, F., Kaur, H., Condiff, K., Chang, S. and Terveen, L. 2015. Putting Users in Control of Their Recommendations. *Proc. ACM RecSys '15*, 3–10.
- [9] Hill, W., Stead, L., Rosenstein, M. and Furnas, G. 1995. Recommending and evaluating choices in a virtual community of use. *Proc. CHI '95*, 194–201.
- [10] Kapoor, K., Kumar, V., Terveen, L., Konstan, J.A. and Schrater, P. 2015. "I Like to Explore Sometimes": Adapting to Dynamic User Novelty Preferences. *Proc. ACM RecSys '15*, 19–26.
- [11] Kaptein, M., Markopoulos, P., de Ruyter, B. and Aarts, E. 2015. Personalizing persuasive technologies: Explicit and implicit personalization using persuasion profiles. *International Journal of Human-Computer Studies*. 77, (May 2015), 38–51.
- [12] Knijnenburg, B., Willemsen, M., Gantner, Z., Soncu, H. and Newell, C. 2012. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction*. 22, 4–5 (Oct. 2012), 441–504.
- [13] McNee, S., Kapoor, N. and Konstan, J.A. 2006. Don't Look Stupid: Avoiding Pitfalls When Recommending Research Papers. *Proc. ACM CSCW '06*, 171.
- [14] McNee, S., Riedl, J. and Konstan, J.A. 2006. Making recommendations better: an analytic model for human-recommender interaction. *CHI '06 Extended Abstracts* (2006), 1103–1108.
- [15] Pariser, E. 2011. *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. Penguin.
- [16] Phelan, O., McCarthy, K., Bennett, M. and Smyth, B. 2011. Terms of a Feather: Content-Based News Recommendation and Discovery Using Twitter. In *Advances in Information Retrieval*, LNCS 6611. Springer. 448–459.
- [17] Prochaska, J.O. and DiClemente, C.C. 1984. *The transtheoretical approach: crossing traditional boundaries of therapy*. Dow Jones-Irwin.
- [18] Prochaska, J.O. and Norcross, J.C. 2001. Stages of change. *Psychotherapy: Theory, Research, Practice, Training*. 38, 4 (2001), 443–448.
- [19] Rutjes, H., Willemsen, M.C. and IJsselstein, W.A. 2016. Understanding Effective Coaching on Healthy Lifestyle by Combining Theory- and Data-Driven Approaches. *Proc. Workshop on Personalization in Persuasive Technology at Persuasive 2016*.
- [20] Starke, A., Willemsen, M.C. and Snijders, C. 2015. Saving Energy in 1-D: Tailoring Energy-saving Advice Using a Rasch-based Energy Recommender System. *Proc. 2nd Int. Workshop on Decision Making and Rec. Sys.*, 5–8.
- [21] Suchman, L. 1993. Foreword. *Participatory Design: Principles and Practices*. D. Schuler and A. Namioka, eds. CRC / Lawrence Erlbaum Associates. vii–x.
- [22] Ziegler, C.-N., McNee, S., Konstan, J.A. and Lausen, G. 2005. Improving Recommendation Lists through Topic Diversification. *Proc. WWW '05*, 22–32.