"Let me tell you who you are"

Explaining recommender systems by opening black box user profiles

David Graus, Maya Sappelli, Dung Manh Chu
FD Mediagroep
Prins Bernhardlaan 173, 1097BL
Amsterdam, The Netherlands
{firstname}.{lastname}@fdmediagroep.nl

ABSTRACT

Personalization of media services is gaining more and more traction, e.g., through the rise of personalization driven by recommender systems across media outlets. At the same time, we see a general rise in distrust and skepticism around the collection and processing of personal data, spurred by policy changes such as the introduction of the GDPR, data breach incidents, and the rise of privacy concerns in general. We feel it is of central importance to be transparent about the information we collect, and the ways we use it. In this position paper we motivate the importance of enabling transparency through explaining our recommender system. More specifically, we aim to explain the inferred user profiles that are central to content-based recommender systems. We describe how user profile explanations can contribute to, or enable different aspects of our recommender system; *transparency* to help users better understand the inner workings of the recommender system, *scrutability* to allow users to provide explicit feedback on the internally constructed user profiles, and *self-actualization* to support users in understanding and exploring their personal preferences. Finally, we believe that user profile explanations can find novel and interesting explanations as an end in itself.

ACM Reference Format:
David Graus, Maya Sappelli, Dung Manh Chu. 2018. "Let me tell you who you are": Explaining recommender systems by opening black box user profiles. In *FATREC 2018 Workshop: Responsible Recommendation.*

1 INTRODUCTION

Personalized experiences powered by recommender systems have, after years of being a mostly academic endeavor, finally permeated our daily lives. Whether it is through personalized recommendations in web-shops (e.g., Amazon), personalized media consumption (e.g., Spotify, Netflix), search engines (e.g., Google), or virtual assistants (e.g., Apple's Siri, Microsoft's Cortana).

However, driven by data breach incidents and ad-driven business models, at the same time we're seeing a rise in distrust and skepticism around the collection of personal data—a requirement for recommender systems. In addition, the EU's GDPR has generated an increased interest in aspects of explainability and transparency of black-box machine learning algorithms and models.

The FD Mediagroep (FDMG[1]) is the primary source of financial economic news in the Netherlands, through "Het Financieele Dagblad" (FD) a daily financial newspaper (similar in nature to *The Financial Times*), and the only commercial news radio station in the Netherlands ("BNR Nieuwsradio"). Serving a wide variety of users with different backgrounds, interests, and contexts, we believe that serving our digital media (text and audio) in a personalized manner can be beneficial. In the light of this, we are working on two Google DNI[2]-funded projects that involve (a) enabling a novel non-linear radio experience at BNR, through automatically generating personalized playlists that match our listener's interests (*SMART Radio*), and (b) personalized summaries of articles to better match the preferences, interests, and profiles of our users (*SMART Journalism*) [10].

2 EXPLAINING RECOMMENDATIONS

Personalization requires the collection of users' consumption behavior (e.g., reading or listening behavior), to estimate preferences and interests, and better serve users more relevant content. In the wake of the aforementioned developments such as the GDPR and the general increased awareness around processing and storing data, the recommender systems community recently focuses strongly on topics such as transparency and explainability.

Explaining recommendations can serve different purposes [12]. Early efforts focus on explaining how the system works (i.e., *transparency*), e.g., by exposing structured representations of user profiles [6]. More recent efforts tend to also focus on *increasing the effectiveness* of recommender systems, and on helping users in taking decisions, e.g., through showing a user why an item matches their profile [9, 11]. Furthermore, explaining recommendations can enable or improve the *scrutability* of a recommender system, i.e., by allowing users to tell that the system is wrong. Overall it has been found that explanations of recommender system decisions may increase *trust*, by increasing a user's confidence in the system [1, 12].

2.1 Content-based recommender systems

Before we detail our planned approach of user profile explanations, we first describe the (slightly simplified) basic methodology of content-based recommender systems.

First, we model the task of generating a recommendation as scoring an item i's relevance for a user profile u. A recommender system thus takes as input a user profile (u), and a set of (candidate) items ($i \in I$). Items are represented as (high-dimensional) feature vectors, i.e., $i = [f_1, f_2, f_3, \ldots, f_n]$ where f corresponds to a feature that represents an aspect or property of an item, e.g., descriptive metadata such as authors, sections (e.g., in a one-hot-encoded representation), or content representations, such as topics, words, etc.

[1] http://www.fdmediagroup.com

[2] https://newsinitiative.withgoogle.com/dnifund/

,
.

Typically, user profiles u are constructed from aggregating feature vectors that represent items a user has consumed [8], e.g., by taking statistics from the aggregated features, such as means and standard deviations. Then, a scoring function (F) is learned to compute a score given the user profile u and item i, i.e. $S = F(u, i)$.

3 EXPLAINING USER PROFILES

In our *SMART Journalism* project, we aim to construct user profiles from aggregating features of consumed items, e.g., content representations such as topics, word embeddings, and entities, but also descriptive metadata such as authors, sections, or added tags [10], i.e., they will be represented as high dimensional feature vectors and used in a learning to rank (LTR)-based recommender system [3]. We posit that explaining the typically "black box" user profiles, that serve as the recommender system's input can be beneficial in several aspects, which we further detail below.

3.1 Explaining Input, Not Output

Typical content-based recommender explanations are item-based, and focus on explaining why a specific item has been recommended [2]. In essence, these methods inform the user of the item feature(s) that contribute to a high matching score, and explain the recommender system's *output* (i.e., the item i that got the highest score S).

By exposing the user profile we can support users in better understanding part of the underlying mechanisms of our recommender system: the profile corresponds to (half of) the recommender system's *input*, i.e., the user profile u (the other half being a candidate item i). Item-based explanations on the other hand explain only a small fraction of the system's *output*, typically a subset of aspects of a single item i.

With item-based explanations, a user is typically offered (a small subset of) item features of a recommended item i, which leaves them to infer aspects of its own unique user profile u (e.g., with each recommended item, one or more aspects of u may be exposed). Switching the focus to explaining the system's input (i.e., user profile u), enables a more complete picture, as the number of aspects to expose is not restricted by the set of recommended items. It hence allows to better explain the inner workings of the recommender system, and thus more effectively increase *transparency*.

Since our user profiles will be high dimensional, exposing the complete profile may be infeasible. Our main challenge is to develop a method for effectively selecting which aspects to explain, e.g., focusing on abnormalities [7] by measuring which features of a profile deviate from the (global) average. More generally, our goal is to summarize and visualize high dimensional data, to effectively expose (aspects of) user profiles in such a way that users can interpret them, and take action on them.

3.2 Scrutability

To increase the scrutability of the recommender system [12], we may elicit user feedback on the user profiles, e.g., by providing users the ability to correct their profiles when they disagree with (parts of) it. In doing so, we enable scrutability of the system's input (the profile), whereas enabling scrutability of the system's output (a recommendation) is more commonly seen with content-based recommender systems [13].

Such an "explicit" feedback signal can find multiple applications; First, this feedback can be employed for directly adjusting user profiles, e.g., through discounting user profile aspects/features that a user disagrees with, or vice versa, by adding aspects the system may have not (yet) picked up from reading behavior. Second, this feedback may be employed for improving the recommender system's accuracy. In conjunction with implicit feedback (e.g., clicks and skips)—typically used in learning to rank-based recommender systems [5]—a user's explicit feedback can prove a valuable additional signal for training or evaluation purposes.

3.3 Self-Actualization

Self-actualization concerns itself with supporting users in developing, exploring, and understanding their unique personal preferences [4]. User profiles may offer readers insights into their own behavior, biases, interests, or expertise, and effectively exposing a user profile can help users who actively seek to reduce biases to do so. More generally, eliciting exploration may yield benefits such as unleashing long tail articles, and more generally increasing user engagement through broader and more diverse consumption of content.

4 CONCLUSIONS

In summary, in this position paper we described how explaining user profiles in content-based recommendation can serve to increase transparency, scrutability, and enable self-actualization.

REFERENCES

[1] Jonathan L. Herlocker, Joseph A. Konstan, and John Riedl. 2000. Explaining Collaborative Filtering Recommendations. In *Conference on Computer Supported Cooperative Work (CSCW '00)*. ACM, New York, NY, USA, 241–250. https://doi.org/10.1145/358916.358995

[2] Yunfeng Hou, Ning Yang, Yi Wu, and Philip S. Yu. 2018. Explainable recommendation with fusion of aspect information. *World Wide Web* (13 4 2018), 1–20. https://doi.org/10.1007/s11280-018-0558-1

[3] Alexandros Karatzoglou, Linas Baltrunas, and Yue Shi. 2013. Learning to Rank for Recommender Systems. In *RecSys (RecSys '13)*. ACM, New York, NY, USA, 493–494. https://doi.org/10.1145/2507157.2508063

[4] Bart P. Knijnenburg, Saadhika Sivakumar, and Daricia Wilkinson. 2016. Recommender Systems for Self-Actualization. In *RecSys (RecSys '16)*. ACM, New York, NY, USA, 11–14. https://doi.org/10.1145/2959100.2959189

[5] Jiahui Liu, Elin Pedersen, and Peter Dolan. 2010. Personalized News Recommendation Based on Click Behavior. In *IUI*.

[6] Stuart E. Middleton, Nigel R. Shadbolt, and David C. De Roure. 2004. Ontological User Profiling in Recommender Systems. *ACM TOIS* 22, 1 (Jan. 2004), 54–88. https://doi.org/10.1145/963770.963773

[7] Christoph Molnar. 2018. *Interpretable Machine Learning*. https://christophm.github.io/interpretable-ml-book/. https://christophm.github.io/interpretable-ml-book/.

[8] Michael J. Pazzani and Daniel Billsus. 2007. *Content-Based Recommendation Systems*. Springer Berlin Heidelberg, Berlin, Heidelberg, 325–341. https://doi.org/10.1007/978-3-540-72079-9_10

[9] Bashir Rastegarpanah, Mark Crovella, and Krishna P. Gummadi. 2017. Position Paper: Exploring Explanations for Matrix Factorization Recommender Systems. In *FATREC '17*.

[10] Maya Sappelli, Dung Manh Chu, Bahadir Cambel, David Graus, and Philippe Bressers. 2018. SMART Journalism: Personalizing, Summarizing, and Recommending Financial Economic News. In *The Algorithmic Personalization and News (APEN18) Workshop at ICWSM '18*.

[11] Maartje ter Hoeve, Mathieu Heruer, Daan Odijk, Anne Schuth, Martijn Spiers, and Maarten de Rijke. 2017. Do News Consumers Want Explanations for Personalized News Rankings?. In *FATREC '17*.

[12] Nava Tintarev. 2010. *Explaining recommendations*. Ph.D. Dissertation. University of Aberdeen.

[13] Yongfeng Zhang and Xu Chen. 2018. Explainable Recommendation: A Survey and New Perspectives. *CoRR* abs/1804.11192 (2018). arXiv:1804.11192 http://arxiv.org/abs/1804.11192