UvA-DARE (Digital Academic Repository)

3bij3: A framework for testing effects of recommender systems on news exposure

Löcherbach, F.; Trilling, D.

Link to publication

3bij3: A framework for testing effects of recommender systems on news exposure

Felicia Löcherbach
Department of Communication Science
Vrije Universiteit Amsterdam
Amsterdam, Netherlands
f.loecherbach@vu.nl

Damian Trilling
Department of Communication Science
University of Amsterdam
Amsterdam, Netherlands
d.c.trilling@uva.nl

Abstract—We developed 3bij3, a framework that presents a news app to participants, with contents that are displayed based on different recommendation logics. It tracks usage over time and enables large-scale field experiments.

Index Terms—recommender systems, news exposure, communication science, web application

I. Introduction

The impact of recommender systems on how we consume and perceive news is still understudied. To shed more light on this, we developed one of the first software solutions for such studies, and present *3bij3*. 3bij3 means three by three in Dutch, and signifies the most prominent feature of a news application that forms the interface for study participants. It displays a 3x3 grid of the nine most relevant news articles to a user, selected by different mechanisms (e.g., explicit customization vs based on past behaviour vs random). 3bij3 can be used to conduct large-scale field experiments, in which participants' use of the site can be tracked over extended periods of time.

II. Related Research

Social-scientific experiments [1]–[4], usually based on the creation of a set of fictitious news articles regarding a specific topic that are varied on factors such as political stance, often create a highly artificial situation. They thus often lack external and ecological validity. *Input-output analyses* [5], [6] aim at uncovering the workings and effects of existing recommendation algorithms – however, they have to treat the actual algorithm as a black box. Lastly, *simulations* [7]–[10] evaluate recommendations based on existing or simulated datasets using various metrics. While being a good approach for judging the performance of algorithms regarding certain predefined measures, such studies remain in an "evaluation setting where recommendation approaches are compared without user interaction" [11, p.9].

III. 3bij3

Compared to previous work, 3bij3 gives researchers control over the recommendation system under study *and* creates a realistic environment for the participants. Figure 1 gives an

overview of the components of 3bij3: (1) Content retrieval, processing, and enrichment; (2) article selection; (3) user interaction. It integrates web scraping/parsing, supervised machine learning for labeling articles, different recommender systems, a web interface for participants, and gamification elements (earning points for interaction).
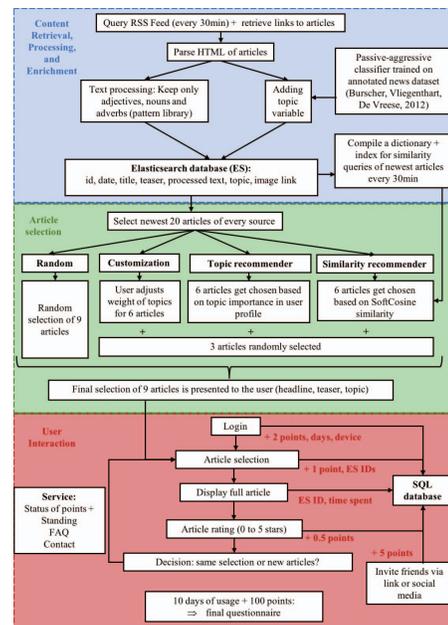
Fig. 1. Overview of the framework

3bij3 is written in Python and uses Flask for the web frontend. When signing up, participants are randomly assigned to one of the article selection conditions. Articles are stored in ElasticSearch, user data (incl. clickstream) in MySQL.

IV. Limitations and future work

3bij3 has been tested on a first group of participants ($N = 25$). We plan to use it in a large-scale study ($N > 500$) in order to test participants' reactions towards different selection logics.

IEEE computer society

REFERENCES

[1] M. A. Beam, "Automating the News: How Personalized News Recommender System Design Choices Impact News Reception," *Communication Research*, vol. 41, no. 8, pp. 1019–1041, 2014.

[2] I. Dylko, I. Dolgov, W. Hoffman, N. Eckhart, M. Molina, and O. Aaziz, "The dark side of technology: An experimental investigation of the influence of customizability technology on online political selective exposure," *Computers in Human Behavior*, vol. 73, pp. 181–190, 2017.

[3] J. A. Yang, "Effects of popularity-based news recommendations (most-viewed) on users' exposure to online news," *Media Psychology*, vol. 19, no. 2, pp. 243–271, 2016.

[4] R. K. Garrett, "Echo chambers online?: Politically motivated selective exposure among Internet news users," *Journal of Computer-Mediated Communication*, vol. 14, no. 2, pp. 265–285, 2009.

[5] D. O'Callaghan, D. Greene, M. Conway, J. Carthy, and P. Cunningham, "Down the (White) Rabbit Hole: The Extreme Right and Online Recommender Systems," *Social Science Computer Review*, vol. 33, no. 4, pp. 459–478, 2015.

[6] M. Haim, A. Graefe, and H. B. Brosius, "Burst of the Filter Bubble?: Effects of personalization on the diversity of Google News," *Digital Journalism*, vol. 6, no. 3, pp. 330–343, 2018.

[7] M. Sood and H. Kaur, "Survey on News Recommendation," *International Journal of Advanced Research in Electrical*, vol. 3, no. 6, 2014.

[8] J. Möller, D. Trilling, N. Helberger, and B. van Es, "Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity," *Information Communication and Society*, pp. 1–19, 2018.

[9] D. Bridge and J. P. Kelly, "Ways of computing diverse collaborative recommendations," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 4018 LNCS. Springer, 2006, Conference Proceedings, pp. 41–50.

[10] M. Ö. Karakaya and T. Aytekin, "Effective methods for increasing aggregate diversity in recommender systems," in *Knowledge and Information Systems*, 2017, pp. 1–18.

[11] M. T. Ribeiro, N. Ziviani, E. S. D. Moura, I. Hata, A. Lacerda, and A. Veloso, "Multiobjective Pareto-Efficient Approaches for Recommender Systems," *ACM Transactions on Intelligent Systems and Technology*, vol. 5, no. 4, pp. 1–20, 2014.