Report Paper for "Exploring the Filter Bubble: The Effect of Using Recommender Systems on Content Delivery" by Nguyen, Hui, Harper, Terveen and Konstan

Seminar: Big Data, AI and Society
University of Zuerich
Philip Hofmann, 14-710-842

*** Summary of the paper ***

The authors of this paper researched the question, if recommender systems in general would lessen the content diversity of consumed items by a user. For this study, they looked at the movie recommender system used by the MovieLens platform. After analyzing data of users of this platform, specifically movies that have been recommended to these users and the rating these users gave to movies, they concluded, that all users suffered from a content narrowing consumption over time, but this effect was actually lessened for the users that followed movie recommendations, which would suggest that the recommender system on this particular platform had the opposite effect of a filter bubble in comparison to a control group.

*** Three ***positive*** aspects of the paper ***

The authors of the paper took many things into consideration, like the cleaning of data analyzed for this study in a temporal manner. Meaning removal of a specific time period in the beginning in case of the users rating already known movies. I also like that the authors did not only look at content diversity over time, but also at an experience grading and then followed up with research question c. Although we have to keep in mind that the main research question is about the content diversity of consumed items and not enjoyability, since this was about the validation of an actual "filter bubble" occurring.

*** Three ***negative*** aspects of the paper ***

While the authors analyzed data gathered through the MovieLens platform, they should have been more careful with the distinction of movies "consumed", since they in reality do not measure consumption of movies, but rating. A careful distinction would have been important, since we have no guarantee that every movie watched will also be rated, and it would be important to catch every movie consumed/watched, since otherwise we cannot have an accurate representation of the affect the recommender system had on the user. Also the 3-hour timelimit given on recommendation and its rating by the authors might be too large, allowing the user to see a recommendation, then watch it, then rate it, and it not being included into the data analysis because it happened in under 3 hours. Additionally, I do not see the used dataset publicly linked or attributed to by the authors by the paper, which makes validation of the study difficult.

*** Evaluation of related work ***

The authors listed similar papers concerning the topic of opinions and recommender systems, showed that those already published papers sometimes have weaknesses and they argue that their paper is the first one in the specific regards of recommender systems leading to a narrowing of consumed content.

*** Detailed evaluation ***

I particularly liked the authors considerations concerning the rating blocks definition, including the incorporation of specific time periods of inclusion or exclusion of data depending on observations of users rating at the beginning movies that they had already consumed in their past history, not being affected by the recommender system of this platform and letting the recommender system gathering data on the user to increase its effectiveness overall.

I am not specifically fond of the 3 hour time exclusion from the recommendation of a movie to its rating, since when observing my own behavior, if I want to watch a movie, I access the internet and scroll through some suggested movies and watch it, this happens within a 3 hour time period and therefore my data points would have been excluded in such a study. It would have seen reasonable to me to gather data on the average movie length of movies recommended by this platform, and then setting a time limit that is well beneath this average time limit, so we do include data points where a movie is immediately consumed and rated after a recommendation, but we do exclude someone rating a movie right after seeing the recommendation because of past consumption.

Like already mentioned, a careful distinction between “consumption” and “rating” must be done. From the methodology, it is clear, that the rating of a movie by a user was measured, and not the consumption. Sentences like “378 accessed ‘Top Picks For You’ but never consumed any recommendation” is misleading and needs to be clarified that the authors so only have knowledge on the effect of the recommender system on movie ratings of this user, but not on consumption of this user.

The grouping of the users into “following” and “ignoring” should have been more fine-grained, by adding a third group. The ignoring group is the biggest group and is a mash-up between users that consume the recommendations but did not rate any of the recommended movies, and users that never saw the recommendations. These groups should be separated, because we know of a surety that the latter of the users mentioned cannot possibly have been affected by the recommender system since they have never seen any recommendation. Also, when considering Figure 4, I am asking myself why the huge middle-group of users between the two extremes, or cut points, have not been considered at all in this study. It is unfeasible to assume a real-world model where all users must show an extreme bias towards a specific behavior.

Even though it was not part of the initial research topic (if filter bubble applies), I liked the inclusion of the experience-question of the two groups into the research. I would have liked to see also the investigation of why a narrowing of content diversity happens over time in both groups, and why exactly this effect was narrowed for the Following-Group, which would need more detailed analysis (how diverse were the recommended movies from the rated movies of the user? If the recommended movies are different in content, we might attribute this effect to the recommender system, but if the recommended movies are all similar in content, this would hint towards external factors like users of a specific group being more explorative in their movie selection, maybe using also other sources of recommendations).

*** Reproducibility ***

The authors of the paper explained and visually represented their methodologies, approaches, data points inclusion or exclusion and formulas used. They also explained which data was collected and used for the study. But I did not see any hints towards if the data would be publicly accessible in the paper. For this study to be reproducible, we would need the whole dataset publicly accessible, including the data points of all users from the specified time-period, especially before data points cleaning (for example full rating of users, without first three months removed).

*** Editorial remarks ***

The equating of rating and consumption is misleading and simply erroneous as used in this paper. They defined the data points collected, and ratings does simply not equate consumption. The conclusion that a group of users never consumed a recommended movie cannot be substantiated with the given dataset used by the authors, the only correct statement would be that this group of users never rated recommended movies.

*** Standard of writing ***

I perceive the paper as well-written and easy to follow. The illustrations distributed in the paper greatly help in the comprehensibility of their methodology used for data analyzation.

*** Overall judgment ***

Weak accept (1). As soon as they remediate the erroneous equation of consumption with rating, it will be a strong accept (2).

*** Statement of confidence ***

high