PrCP: Pre-recommendation Counter-Polarization

Mahsa Badami[1], Olfa Nasraoui[1] and Patrick Shafto[2]
[1] *Knowledge Discovery and Web Mining Lab, Computer Science and Computer Engineering Department, University of Louisville, Louisville, KY, USA*
[2] *Department of Mathematics and Computer Science, Rutgers University - Newark, Newark, NJ, USA*
{*Mahsa.Badami, Olfa.Nasraoui*}*@Louisville.edu, patrick.shafto@gmail.com*

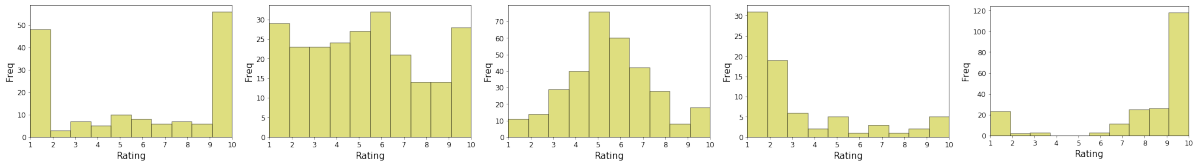
Keywords: Recommender System, Polarization, Controversy, Big data, Algorithmic bias

Abstract: Personalized recommender systems are commonly used to filter information in social media, and recommendations are derived by training machine learning algorithms on these data. It is thus important to understand how machine learning algorithms, especially recommender systems, behave in polarized environments. We investigate how filtering and discovering information are affected by using recommender systems. In the first part of our paper, we study the phenomenon of polarization and its impact on filtering and discovering information. We study polarization within the context of the user's interactions with a space of items and how this affects recommender systems. We then investigate the behavior of machine learning algorithms in environments where polarization emerges, and find that Matrix Factorization models find it easier to learn in polarized environments, and this, in turn, encourages filter bubbles which reinforce polarization. Finally, building on a methodology for quantifying the extent of polarization in a rating dataset, we propose new counter-polarization approaches for existing collaborative filtering recommender systems, focusing particularly on state of the art models based on Matrix Factorization. We propose a new recommendation model for combating over-specialization in polarized environments toward counteracting polarization in human-generated data and machine learning algorithms.

1 Introduction

The growing popularity of online services and social networks and the trend to integrate Recommender Systems (RS) within most e-commerce applications and social media platforms to help filter data to the users, has led to a dynamic interplay between the information that users can discover and the algorithms that filter such information (Melville and Sindhwani, 2011; Bobadilla et al., 2013; Badami et al., 2018; Sun et al., 2018). This has given rise to several side effects, such as algorithmic biases (Dandekar et al., 2013; Baeza-Yates, 2016), filter bubbles (Liao and Fu, 2014), and human-algorithm iterated bias (Shafto and Nasraoui, 2016) and polarization (Morales et al., 2015). Recent research has studied different types of biases generated due to algorithms, including bias and fairness in machine learning (Hardt et al., 2016; Caliskan et al., 2017; Kamishima et al., 2011; Fish et al., 2016); as well as algorithmic bias(Hajian et al., 2016; Baeza-Yates, 2016; Kirkpatrick, 2016; Bozdag, 2013; Lambrecht and Tucker, 2018), and assimilation bias (Zhang et al., 2017). Polarization around controversial issues have arguably affected recommender systems (and vice-versa) (Garimella et al., 2016b; Isenberg, 1986). An effective and efficient recommender system should be able to provide the most suitable recommendation method even in the presence of a set of polarized items. When such issues emerge on social media, we often observe the creation of "echo chambers" or "Filter Bubbles", where there is greater interaction between like-minded people who reinforce each other's opinion (Garimella et al., 2016b). These individuals do not get exposed to the views of the opposing side, and this in turn exacerbates polarization (Dandekar et al., 2013). Allowing users to discover different viewpoints could allow them to develop unique tastes and diverse perspectives (Knijnenburg et al., 2016).

In order to give the users a choice to see more items, we believe that a recommender system should have a systematic mechanism that enables users to discover novel items whose discovery may become hindered as a result of the users' continuous engagement with a system that is continuously learning from this engagement. This is not necessarily the same as recommending a random item by trial and error or by diversifying the recommendation list, to in-



(a) $\Phi_i = 1$ (b) $\Phi_i = 0$ (c) $\Phi_i = 0$ (d) $\Phi_i = 0$ (e) $\Phi_i = 0$

Figure 1: Polarization scores for different rating histograms

crease diversity and serendipity. It is important to note that recommender systems that improve diversity and serendipity are not the same as polarization aware recommender systems. This is because the former generally require diversity in the actual description or nature of items, which in turn requires content data. Our work primarily focuses on items that can cross polarization boundaries, where polarization is based on how users interact with the items (via ratings) and not their content.

Research on polarization in recommender systems has emerged rapidly, in recent years, as an important interdisciplinary topic (Munson and Resnick, 2010; Dandekar et al., 2013; Garimella et al., 2016b; Mejova et al., 2014; Nasraoui and Shafto, 2016; Abisheva et al., 2016; Morales et al., 2015; Matakos and Tsaparas, 2016; Garimella et al., 2016a), with some efforts trying to decrease online polarization, especially in recommender systems (Garimella et al., 2016b; Liao and Fu, 2014; Garimella et al., 2016a; Badami et al., 2017) However, most current work on polarization has either been limited to simple artificial toy problems (two users) (Dandekar et al., 2013), has relied on textual content to detect sentiment and then polarization, or has been confined to specific domains where contentious issues lead to polarization. To date, most content-based studies have been typically conducted within the context of political (or other controversial domain) news and blogs. In this paper, we are more interested in studying the emergence and aggravation of polarization as a result of using collaborative filtering recommender systems.

Aiming toward alleviating the important problems of over-specialization and concentration bias, especially in a polarized environment, and enhancing the usefulness of collaborative filtering RS, we propose a new approach to generating recommendation lists based on a modified Non-Negative Matrix Factorization approach. We formulate theoretically-grounded scenarios for polarization which will allow a simulation-based analysis of the emergence of polarization, as well as designing new counter-polarization strategies for recommender systems. Our proposed approach alters only the input ratings based on the automatically detected polarization of the items and the user's pre-specified tolerance for discovery. The proposed model aims to achieve a trade-off between accurate personalized recommendations and expanding the space of items that can be discovered, hence escaping a filter bubble. Whether humans prefer to discover more or less is beyond the scope of this paper. The proposed pre-recommendation approach is useful for other applications where a dataset should be either published by an online recommender system provider or by researchers. In addition, we propose an Interactive Recommender System (IRS) inspired by (Dandekar et al., 2013) to assess the effect of the proposed strategy on the diversity of recommendations in a polarized environment. We see the proposed simulation approach as a complementary method to investigate the performance of a recommendation process in a polarized environment in an offline experimental setting.

The remainder of this paper is organized as follows. Section 2 presents our proposed methods for handling polarization in recommender systems, followed by experiments in Section 3. Finally, we make our conclusions in Section 4.

2 Proposed Method

In this section, we propose a strategy that can counteract population polarization, independent of a RS algorithm. This means that it can later be employed in a pre-filtering stage along with any recommender system algorithm. Our proposed approach can be used to handle polarization without compromising too much on relevance-based (i.e. pure rating) predictive accuracy. This is a useful strategy since most online system providers are using a RS as a black box; hence, it is difficult to look into the inner workings of the algorithm to modify it.

One might also think of alternatives such as a straightforward counter-polarization approach, consisting of just including some randomly selected items from the opposite view. Temporarily, this would seem to solve the filter bubble problem and increase the diversity of the recommended list. However it would cause much information loss which leads to recom-

mending irrelevant items and eventually risks reducing user satisfaction with the system. In addition, such a remedy is not able to solve the filter bubble problem for a long period.

Finally, our proposed approach works in the context of the classical collaborative filtering (CF) Recommender system algorithm, however, unlike these recommender systems, our proposed algorithm allows each user to *control* how much information to see from opposite views. Similar to CF recommender systems, we also use latent factor models, specifically Non-negative Matrix (NMF), to characterize both items and users based on a set of factors inferred from user-item rating patterns. However, the proposed approach is not specific to NMF and can be easily extended to any RS method. The goal of our proposed recommender system is to avoid guiding the user toward the most popular items and rather to include items that help users become aware of other items that they are not able to discover on their own.

2.1 Problem Definition

We start with our definition of polarization and then define the problem of polarization-aware collaborative filtering (CF).
In the absence of polarization, the distribution of opinions is either J-shaped as in figure 1d and figure 1e, or bell shaped, as in figure 1c. However, as polarization emerges, the resulting distribution shifts to a U-shaped distribution, see figure 1a, with two peaks emerging around the two dominant and confronted opinions at the extreme sides of the rating scale (Badami et al., 2017; Morales et al., 2015; Matakos and Tsaparas, 2016). There are some cases with a flat distribution, 1-b, which represent diverse opinions toward an item. Different examples of such distributions are shown in figure 1 [1].

Definition 1 - Polarization:
Given an environment $G = (U, I, R)$, user $u \in \mathbb{R}^{1\times n}$ had rated item $i \in \mathbb{R}^{m\times 1}$ with rating $r_{ui} \in \mathbb{R}^{m\times n}$ on a scale of x to y. Item i's polarization score ϕ_i is a measure that captures the presence of a gap between opposite concentration poles or opposite polarity peaks of the histogram of all users' ratings r_i, when such poles exist. We say the item is polarized if $\phi_i \geq \delta$. This definition is subjective and tries to define an intuitive but data-driven notion of polarization. Instead of using an explicit formula to calculate this score, we compute it using a data driven machine learning model that learns to automatically assign a polarization score to any rating histogram after training the model on real item rating histograms that have been manually labeled according to their polarization level (Badami et al., 2017)

Definition 2 - Polarization-aware collaborative filtering recommendation:
Given a set of ratings $R \in \mathbb{R}^{m\times n}$ collected from a set of users $U \in \mathbb{R}^{1\times n}$ for a set of items $I \in \mathbb{R}^{m\times 1}$, the problem of polarization-aware collaborative filtering recommendation (CF) can be modeled by the triplet (U, I, R), in a way that a recommender system should recommend a ranked item set $i_1, ..., i_t \in I$ according to 1) the relevance of the item to the user's interest, and 2) the item's polarization score. As a realization from definition 2, (U, I, R) can be denoted by (u, i, r) which means that user u rated item i with value r.

2.2 Pre-recommendation: Countering Polarization (PrCP)

In this step, we aim to transform the source data in such a way that it mitigates extreme ratings that make an item polarized. By doing this, we still keep the user's relative preferences, yet make it more moderate so that no extreme recommendation can be generated from a standard recommender system algorithm. We perform a controlled distortion of the training data based on which a recommender system is trained to help the users receive more useful recommendations, in the presence of polarization. This transformation is based both on the user's willingness to discover more items and on the item's polarization score.

The proposed solution to counteract polarization by making the training dataset less polarized, employs a stochastic mapping function as defined below:

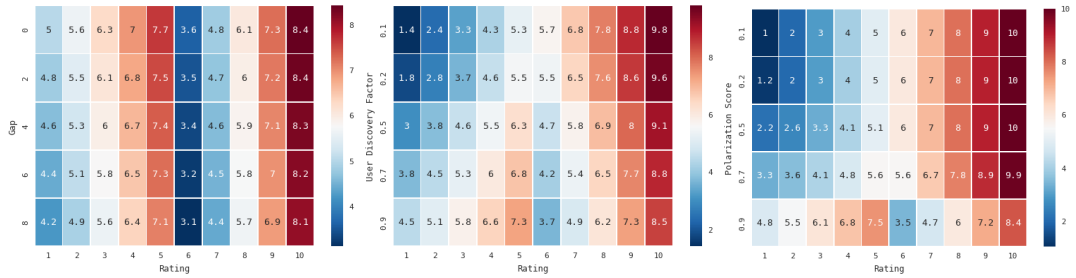
$$f : (U, I, R) \rightarrow (U, I, R') \text{ with probability } p \qquad (1)$$

The function transforms a user-item rating, r_{ui} (for user u on item i) into rating r'_{ui}, based on the rating itself, population average rating, item's polarization score and user's chosen discovery factor, as follows.

$$\begin{aligned} r'_{ui} &= r_{ui} - \lambda_u \times (\bar{r} + \frac{g_i}{g_{max}}) \times \Phi_i^{\lambda_u + r_{ui}} \quad \text{if } r_{ui} \geq \delta \\ r'_{ui} &= r_{ui} + \lambda_u \times (\bar{r} - \frac{g_i}{g_{max}}) \times \Phi_i^{\lambda_u + r_{ui}} \quad \text{if } r_{ui} < \delta \end{aligned} \qquad (2)$$

where $\lambda_u \in [0, 1]$ is the user's selected discovery factor. At one extreme, it is 1 when the user indicates that s/he is interested in discovering more items, especially from the opposite view. At the opposite extreme, if the user sets $\lambda = 0$, the result reduces to using only the classical recommendation algorithm which aims to minimize the squared error on the set of raw ratings. Note that if a user expresses an interest in

[1] Each distribution belongs to a movie crawled from IMDb by (Badami et al., 2017) with polarization score, ϕ, calculated by the method presented in the paper.



(a) g and r_{ui} (b) λ and r_{ui} (c) Φ and r_{ui}

Figure 2: Correlation between user discovery factor (λ), polarization score (Φ), rating (r_{ui}) and gap (g) in the pre-recommendation style counter-polarization approach

considering items from the opposite view, it does not necessary mean that s/he would definitely like or purchase those items. The goal here, is to simply give an option to the users to be able to burst out of their filter bubbles. $\Phi_i \in [0,1]$ is the polarization score which is computed using the Polarization Detection Classifier (Badami et al., 2017). $g_i \in [0,1]$ indicates the gap between the two rating extreme ranges for a polarized item, in other words it measures how polarized the user population's ratings are for item i. We define the gap g_i as the difference between an item's typical minimum rating when it is liked and its typical maximum rating when it is disliked. In other words, the gap g_i captures the difference between extreme opinions regarding an item. We define g_i as

$$g_i = \frac{\max_{u \in Liked(i)}(r_{ui}) - \min_{u \in Unliked(i)}(r_{ui})}{\max_u(R_u) - \min_u(R_u)} \quad (3)$$

where $Liked(i)$ is the set of users who liked item(i) (i.e. $r_{ui} \geq \delta$) and $Unliked(i)$ is the set of users who didn't like item(i) (i.e. $r_{ui} < \delta$).

Note that the denominator normalizes the gap by the extremes of the population ratings. g_{max} is simply the difference between the maximum and minimum rating that a typical user can provide for any item, using the system's rating scale. The more polarized a population gets, the higher g_i gets. δ indicates which ratings are considered as liked versus disliked.

3 Experiments

In order to evaluate the impact of our proposed counter-polarization approach, we need to measure the increase in the number of the items from the opposite view that are ever recommended to the user. This is different from catalog coverage, which considers how many of the recommended items belong to the "long tail" of items. In this section we will take a deeper look at the view space coverage and effects of polarization on the algorithms.

To empirically validate our proposed pre-recommendation scheme, we first studied how factors λ_u, Φ_i, g_i would affect the mapping function from section 1. Figures 2a-2c show how the difference among extreme values affects the initial rating r_{ij} in a polarized environment if a user u has a high discovery factor λ. In figure 2a, we assume that ratings are on a scale of 1 to 10 and that all items have the same polarization score, $\forall i \in I, \phi_i = 0.9$. As mentioned before, g_i represents the difference between extreme opinions of an item. For example if $g_i = 2$, item i has received two diverging sets of ratings from users. Users who liked this item rated it 10,9,8,7, while those who did not enjoy the item as much had given ratings in the range of 1 to 4. So there is a 2-gap between the given ratings; hence, the item ratings histogram looks like figure 3. Similarly, figure 2b indicates how the transformation affects the initial ratings for an arbitrary item i, where $g_i = 2$ and the user discovery factor λ is 1. Finally, we study the effect of the user's chosen discovery factor on transforming the source data. Here, we assumed that $g_i = 2$ and that the item is polarized, with $\phi_i = 0.9$. As shown in figure 2c, we performed a controlled distortion of the training data from which a recommender system is learned to help the users receive more recommendations in the presence of polarization. By doing this, we still keep the users' preferences, yet make it more moderate so that, less extreme recommendations are generated when using a conventional recommender system algorithm.

3.1 Experimental Settings

Most data publishers provide information regarding the data collection process, yet there are often hidden biases which affect the recommendation process (Badami et al., 2017; Nasraoui and Shafto, 2016; Baeza-Yates, 2016). Hence, we study the effect of po-

larization on recommender systems on multiple users in a fixed environment, inspired by (Dandekar et al., 2013).

We evaluate the performance of our approach in terms of rating prediction accuracy, using the Mean Squared Error (MSE) (Koren et al., 2009). As part of studying polarization, We also define the Opposite View Hit Rate (OVHR) ratio based on the ratio of the number of items from the opposite view to the total number of recommended items. This metric helps us to verify whether an item from the opposite view is among the recommended items. Considering each user, if any of the items from the opposite view is included in the recommendation list, then a *hit* occurred.

3.2 Simulating the Interactive Recommendation Process

We consider the following simple environment: Let $G = (U, I, R)$ be an environment where user $u \in U$ can rate item $i \in I$ with rating $r_{ui} \in R$ on a scale of x to y. The item could be a book, web page, news article, movie, etc. We define a recommender system algorithm as follows:

Definition 3 : *Let the number of users, $|U| = n$ and number of items, $|I| = m$. A recommender system algorithm takes environment G as input along with a user $u \in U$, and outputs a set of items $i_1, ..., i_{k_t} \in I$.* Thus, given an environment G, representing which users have rated which items and a specific user u, a recommender system algorithm's output is a list of items to be recommended to u. We assume that u has to pick only one item from the recommendation list and that s/he then provides a rating r_{ui} for the selected item.

We generate a rating environment with 50 users and 200 items where items are evenly divided in two opposite viewpoint sets that we refer to as red items and blue items. Users are also divided into two groups based on whether they like *Red* or *Blue* items [2]. Each user $u \in U$ rates half of the items of I, in such a way that the rating r_{ui} is greater than δ if s/he likes item i, and less than δ if s/he does not like it. This process forms environment G. We also assume that users are rational and are truly expressing their preferences with ratings on a scale of 1 to 10. For concreteness, we assumed $\delta = 5$. In order to make the environment polarized, we assume that user $u_a \in GroupA$ likes red items more than blue ones, and hence all of his/her ratings for the red items are higher than all of his/her ratings for the blue ones. Similarly, we assume that user $u_b \in GroupB$ likes blue items more than red ones and hence all of his/her ratings for the blue items are higher than all of his/her ratings for the red ones. Finally, we generated environment G with different values of Gap, g and user's discovery factor, λ_u.

In order to understand the Interactive Recommender System (IRS), we start by showing some experiments that illustrate examples of how such a system works in environment G. In all of the examples, we set the number of factors in the latent space, k_f, to 5 and we compute the list of top $k_t = 5$ items to be recommended to each user. The user will give a rating for only one of the selected items and we take this rating value from the true source of ratings, i.e. the ground-truth data. We repeat this procedure 100 times (there are 100 unrated items for each user) to simulate an interactive recommender system scenario. In each iteration, we measure MSE from the training and testing phases. We also keep track of the items to which a user decided to react by providing a rating.

Figure 4 shows traces from the interactive recommendation system for user $u \in GroupA$, which means s/he likes red items more. We generate environment G considering for example that gap g_i of 2 means that $7 \leq r_{ui} \leq 10$ if $u \in GroupA$ likes item i while $1 \leq r_{ui} \leq 4$ if $u \in GroupA$ does not like item i. Figure 3 shows the rating histogram of items and we can clearly see that the difference between the range of the two sub-populations of ratings given to an item is 2. Figure 4, upper row, shows that a classic state of the art recommender system, in our case NMF, is always going to recommend red items, to which the user had previously shown more interest. Although the red items are relevant, the user *Red* is trapped in a filter bubble that does not allow him/her to explore any items from the opposite color/view, at least not before the user has seen all of the *Red* items, the number of which may be enormous in a real life setting. This finding is in line with finding in most of the literature (Lord et al., 1979; Flaxman et al., 2016) including what Dandekar et. al have proved mathematically for an over-simplified theoretical scenario with simpler CF strategies (Dandekar et al., 2013). The second row shows the testing MSE error for user u_a. MSE decreases as the user provides ratings in each iteration; hence, there are fewer unrated items for the user. We repeated the same experiment for user $u_b \in GroupB$ who likes blue items more than red items and we observed the same pattern as user u_a but with Blue items.

Figure 5 shows the results of applying our proposed pre-recommendation counter polarization (PrCP) strategy on the traditional NMF-based algorithm in environment G for user u_a. As we can see,

[2] These labels are purely for the purpose of analysis and they obviously do not affect the recommender system algorithms.

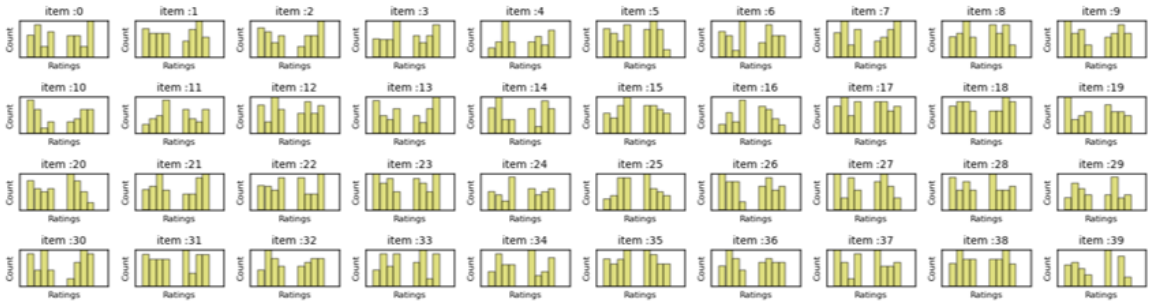


Figure 3: Rating histograms of the items in environment G with polarization ratio 0.25.

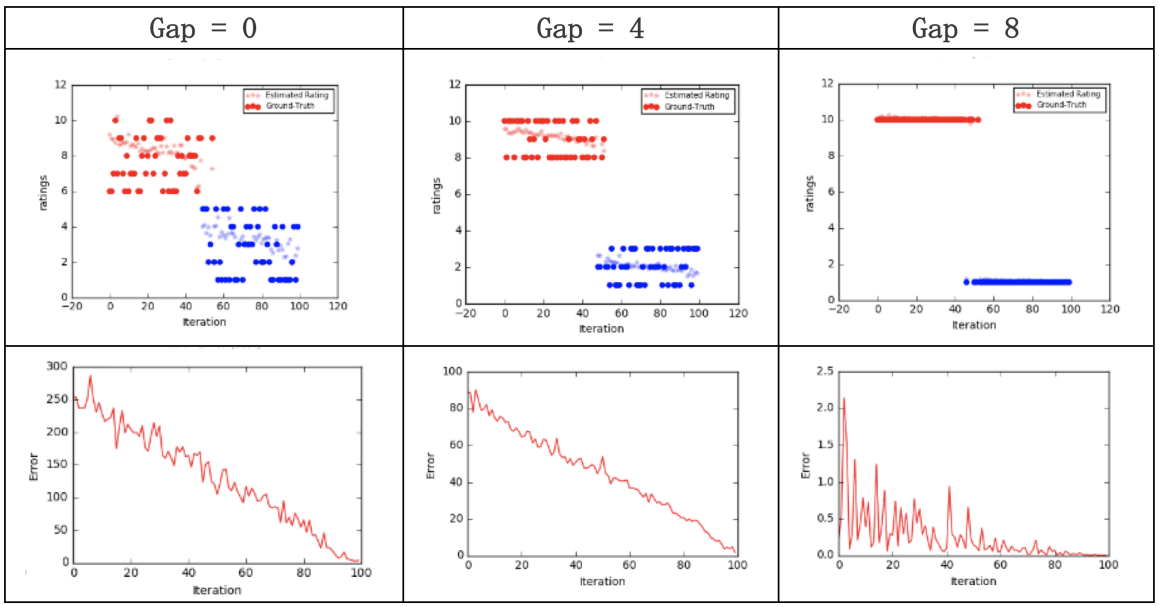


Figure 4: Traces of the Interactive Recommendation process with the classical NMF-based CF recommendation algorithm in environment G with different polarization ratio and gap values, for user u_a who had liked red items more. Although the red items are relevant, the user *Red* is trapped in a filter bubble that does not allow him/her to explore any items from the opposite color/view, at least not before the user has seen all of the *Red* items, the number of which may be enormous in a real life setting.

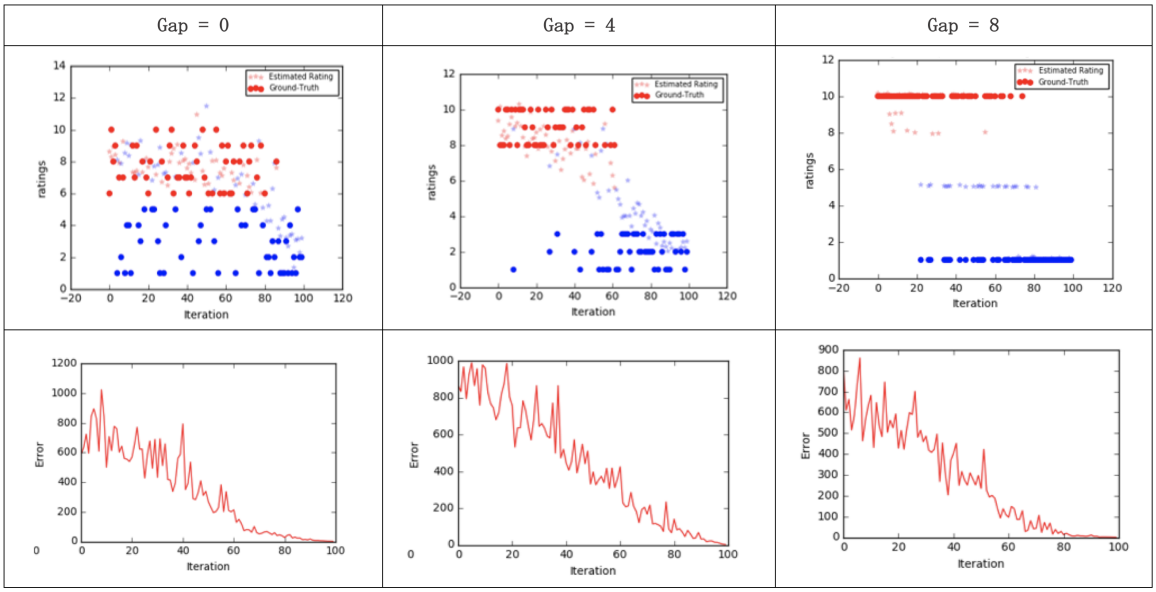


Figure 5: Traces of the Interactive Recommendation process when applying the pre-recommendation counter polarization (PrCP) strategy for user u_a. As we can see, the user gets to see items from different a color/view even in a very polarized environment.

Table 1: Comparison of the counter-polarization methodologies with the classical NMF-based Recommender system in terms of accuracy (on training and testing set, respectively) and opposite view ratio ($OVHR_u$,$OVHR_{k_t}$). There are two scenarios: Scenario (a): same λ for all users and Scenario (b): only user u has $\lambda_u \neq 0$.

			Opposite View Ratio		MSE_{Train}	MSE_{Test}
			$OVHR_u$	$OVHR_{k_t}$		
			mean, std	mean, std	mean, std	mean, std
Classic NMF			0.0 ± 0.00		22.02 ± 5.27	138.96 ± 12.55
PrCP	Scenario (a)	$\lambda_u = 0.2$	$4.8\% \pm 0.06$	$25.0\% \pm 0.035$	126.57 ± 38.13	807.30 ± 70.51
		$\lambda_u = 0.5$	$4.8\% \pm 0.07$	$28.0\% \pm 0.41$	122.38 ± 37.16	805.33 ± 71.77
		$\lambda_u = 1.0$	$5.0\% \pm 0.06$	$2.9\% \pm 0.21$	120.14 ± 34.40	800.23 ± 64.91
	Scenario (b)	$\lambda_u = 0.2$	$5.4\% \pm 0.073$	$4.9\% \pm 0.021$	123.92 ± 36.76	813.01 ± 36.76
		$\lambda_u = 0.5$	$6.2\% \pm 0.075$	$5.2\% \pm 0.042$	122.56 ± 39.081	804.01 ± 75.88
		$\lambda_u = 0.7$	$7.0\% \pm 0.075$	$5.8\% \pm 0.033$	120.97 ± 35.19	803.65 ± 64.65

the user gets to see items from a different color/view even in a very polarized environment. The second row shows the testing MSE error for user u_a which follows the same trend as before since PrCP doesn't change the updating function.

To make a more comprehensive evaluation of performance of the proposed counter-polarization approaches, we repeat the experiment with varying the parameter λ for the proposed counter-polarization methodology. We consider two scenarios: (a) All users have the same λ, i.e. $\lambda_u = c \quad \forall u \in U$, where c is a constant $\in [0,1]$. (b) User u has his/her own unique $\lambda, \lambda_u = c_u$ for user u and $\lambda_u = 0 \quad \forall u \in U - u$, where $c_u \in [0,1]$, is a user defined constant.

The intuition behind this experiment is to study the effect of a user population on recommending items to a single user and to all users. We run the experiments for different $\lambda \in [0.2, 0.7, 1]$ in environment G with $gap = 2$. Then, we compute MSE_{test}, MSE_{train} and $OVHR$ for each user and then take an average over all 50 users. In order to have a comprehensive comparison, we compute $OVHR$ in two ways: (a) $OVHR_u$: Compute the ratio of number of items from the opposite view to what the user has picked from the recommendation list. (b) $OVHR_{k_t}$: the ratio of number of items recommended to the user from an

opposite view.

Table 1 shows that the effects of the two metrics strongly vary depending on the chosen recommendation algorithm and strategy. Trends in varying parameters, show that the higher the user-defined parameter λ, the more she will be recommended items from the opposite view, as desired by the user. When comparing the traditional NMF-based RS with our polarization-aware RS, we see that the traditional NMF-based algorithm achieves good accuracy in rating prediction, yet it is not able to recommend any item from the opposite view. In contrast, Our proposed pre-recommendation scheme can be added to a traditional NMF-based RS and the Polarization-Aware RS would recommend significantly more items ($p \leq 0.05$) from the opposite view compared to the baseline approach, for all the degrees of user-defined discovery factors. These differences between different recommendation processes would go unnoticed if only accuracy measures were considered.

In addition, table 1 shows that having the same user discovery factor for all users has less effect compared to increasing the user discovery factor for a specific user. As we can see in *Scenario (b)*, having an enthusiastic population does not always result in counter-polarization. This effect is even more severe in the polarization-aware strategy where the users do not see any item from the opposite view even when the user population has $\lambda_u = 0.5 \quad \forall u \in U$.

Finally, by looking at the number of recommended items over time in figure 5, we can see that the proposed counter-polarization methodology succeeds to cover items from the opposite view after a few iterations and broadens the viewpoint spectrum even faster if the user is more interested in discovering items from different viewpoints.

4 Conclusions

In this paper, we investigated the mechanism of filtering and discovering information while using recommender systems. We found that environments with different polarization degrees engender different patterns. We proposed a counter-polarization methodology that succeeds to cover items from the opposite view after a few iterations and can broaden the viewpoint spectrum even faster if the user is more interested in discovering items from different viewpoints. The ability of the user to tune the degree of discovery into the opposite viewpoint is an important feature in a polarization-aware recommender system because it allows the users to make decisions about their exploration space. This also contributes to the transparency of a RS algorithm.

5 Acknowledgements

This research was supported by NSF Grant NSF IIS-1549981.

REFERENCES

Abisheva, A., Garcia, D., and Schweitzer, F. (2016). When the filter bubble bursts: collective evaluation dynamics in online communities. In *Proceedings of the 8th ACM Conference on Web Science*, pages 307–308. ACM.

Badami, M., Nasraoui, O., Sun, W., and Shafto, P. (2017). Detecting polarization in ratings: An automated pipeline and a preliminary quantification on several benchmark data sets. In *Big Data (Big Data), 2017 IEEE International Conference on*. IEEE.

Badami, M., Tafazzoli, F., and Nasraoui, O. (2018). A case study for intelligent event recommendation. *International Journal of Data Science and Analytics*, pages 1–20.

Baeza-Yates, R. (2016). Data and algorithmic bias in the web. In *Proceedings of the 8th ACM Conference on Web Science*, pages 1–1. ACM.

Bobadilla, J., Ortega, F., Hernando, A., and Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-based systems*, 46:109–132.

Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and information technology*, 15(3):209–227.

Caliskan, A., Bryson, J. J., and Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334):183–186.

Dandekar, P., Goel, A., and Lee, D. T. (2013). Biased assimilation, homophily, and the dynamics of polarization. *Proceedings of the National Academy of Sciences*, 110(15):5791–5796.

Fish, B., Kun, J., and Lelkes, Á. D. (2016). A confidence-based approach for balancing fairness and accuracy. In *Proceedings of the 2016 SIAM International Conference on Data Mining*, pages 144–152. SIAM.

Flaxman, S., Goel, S., and Rao, J. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, page nfw006.

Garimella, K., De Francisci Morales, G., Gionis, A., and Mathioudakis, M. (2016a). Quantifying controversy in social media. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, pages 33–42. ACM.

Garimella, K., Morales, G. D. F., Gionis, A., and Mathioudakis, M. (2016b). Balancing opposing views to reduce controversy. *arXiv preprint arXiv:1611.00172*.

Hajian, S., Bonchi, F., and Castillo, C. (2016). Algorithmic bias: From discrimination discovery to fairness-aware data mining. In *Proceedings of the 22nd ACM*

SIGKDD international conference on knowledge discovery and data mining, pages 2125–2126. ACM.

Hardt, M., Price, E., Srebro, N., et al. (2016). Equality of opportunity in supervised learning. In *Advances in neural information processing systems*, pages 3315–3323.

Isenberg, D. J. (1986). Group polarization: A critical review and meta-analysis. *Journal of personality and social psychology*, 50(6):1141.

Kamishima, T., Akaho, S., and Sakuma, J. (2011). Fairness-aware learning through regularization approach. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*, pages 643–650. IEEE.

Kirkpatrick, K. (2016). Battling algorithmic bias: How do we ensure algorithms treat us fairly? *Communications of the ACM*, 59(10):16–17.

Knijnenburg, B. P., Sivakumar, S., and Wilkinson, D. (2016). Recommender systems for self-actualization. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 11–14. ACM.

Koren, Y., Bell, R., and Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8).

Lambrecht, A. and Tucker, C. E. (2018). Algorithmic bias? an empirical study into apparent gender-based discrimination in the display of stem career ads.

Liao, Q. V. and Fu, W.-T. (2014). Can you hear me now?: mitigating the echo chamber effect by source position indicators. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 184–196. ACM.

Lord, C. G., Ross, L., and Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of personality and social psychology*, 37(11):2098.

Matakos, A. and Tsaparas, P. (2016). Temporal mechanisms of polarization in online reviews. In *Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on*, pages 529–532. IEEE.

Mejova, Y., Zhang, A. X., Diakopoulos, N., and Castillo, C. (2014). Controversy and sentiment in online news. *arXiv preprint arXiv:1409.8152*.

Melville, P. and Sindhwani, V. (2011). Recommender systems. In *Encyclopedia of machine learning*, pages 829–838. Springer.

Morales, A., Borondo, J., Losada, J. C., and Benito, R. M. (2015). Measuring political polarization: Twitter shows the two sides of venezuela. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(3):033114.

Munson, S. A. and Resnick, P. (2010). Presenting diverse political opinions: how and how much. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1457–1466. ACM.

Nasraoui, O. and Shafto, P. (2016). Human-algorithm interaction biases in the big data cycle: A markov chain iterated learning framework. *CoRR*, abs/1608.07895.

Shafto, P. and Nasraoui, O. (2016). Human-recommender systems: From benchmark data to benchmark cognitive models. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 127–130. ACM.

Sun, W., Nasraoui, O., and shafto, P. (2018). Iterated algorithmic bias in the interactive machine learning process of information filtering. In *Proceedings of the 10th Of the Knowledge Discovery and Information Retrieval conference*.

Zhang, X., Zhao, J., and Lui, J. C. (2017). Modeling the assimilation-contrast effects in online product rating systems: Debiasing and recommendations. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*, RecSys '17, pages 98–106, New York, NY, USA. ACM.