PaRIS: Polarization-aware Recommender Interactive System

Mahsa Badami[1], Olfa Nasraoui[1]

[1]Knowledge Discovery and Web Mining Lab, Computer Engineering and Computer Science Department, University of Louisville, 132 Eastern Parkway, Louisville, Kentucky, USA, 40292

Abstract

One phenomenon that has been recently observed online is the emergence of polarization among users on social networks, where the population gets divided in groups with opposite opinions. As recommender system algorithms become more selective in filtering what users see and discover, one important question arises: Could recommender system algorithms become more selective in filtering what users see and discover? In this paper, we study this question and propose a new counter-polarization approach for existing Matrix Factorization based recommender systems, that can be tuned by a user-controlled counter-polarization parameter which serves like a voluntary user anti-polarization or discovery dial.

Keywords

Recommender System, Polarization, Algorithmic bias, Filter Bubble, Echo Chamber

CEUR-WS.org/Vol-3012/OHARS2021-paper6.pdf

1. Introduction

More than two-thirds of Americans get their news from online social media platforms [1]. At the same time, social media services (e.g. Facebook) automatically filter and sort the order of items on a user's home feed, in many cases without an explicit request from the user to do so [1]. As information filtering algorithms become more and more targeted and selective in filtering what users see and discover, one important question arise: Could these algorithms filter too much information? Despite the numerous benefits of personalized recommender systems, narrow-minded over-specialization, resulting from positive feedback loops, can exacerbate "echo chamber" and "filter bubble" by not expanding, or even restricting, the users' exposure to more diverse and non-obvious options [2, 3, 4, 5]. These two effects get even more extreme in polarized environments leading to an even worse polarized environment [6, 7, 8].

The goal of our work is to expand the domain of recommended options and expand the discovery potential for humans, and thus reducing the users' chance of getting locked in on an online platform. We aim to achieve this goal by designing a recommendation system that not only recommends relevant items but also includes opposite views in case the user is interested in discovering the opposite view. To do so, we propose a novel polarization-aware recommender interactive system (PaRIS) which learns using a modified Non-Negative Matrix

OHARS'21: Second Workshop on Online Misinformation- and Harm-Aware Recommender Systems, October 2, 2021, Amsterdam, Netherlands
Mahsa.Badami@louisville.edu (M. Badami); Olfa.Nasraoui@louisville.edu (O. Nasraoui)
http://webmining.spd.louisville.edu (O. Nasraoui)
© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

[1]https://www.theatlantic.com/technology/archive/2018/04/mark-zuckerberg-atlantic-exclusive/557489/

Factorization objective function. Our counter-polarization approach is modulated by a user-controlled counter-polarization parameter which acts like a user discovery dial to give the user the freedom to tune the severity of their filter bubbles as they wish.

The remainder of this paper is organized as follows. Section 3 presents our proposed methods for handling polarization, followed by experiments in Section 4, and conclusions in Section 5.

2. Related Work

Research on polarization in recommender systems has emerged rapidly, in recent years, as an important interdisciplinary topic [9, 10, 5], with efforts to formulate a richer understanding of the potential characteristics of this Phenomenon and to decrease online polarization, especially in recommender systems [10, 3, 11, 12].

Polarization has been investigated from a network perspective mainly using the social network structure, the content and sentiment of discussions [13, 6], where as others studied polarization based on the *ratings* provided by users on items, within the context of a recommender system [13, 14]. Even though some researchers showed echo chambers in social media are somewhat inevitable by design, other studies proposed strategies to mitigate such effects[15, 16, 17]. Badami et al. proposed a counter-polarization methodology for combating over-specialization in polarized environments [16]. In another research, Tintarev el al. used visual explanations, i.e., chord diagrams and bar charts [18] to address polarization.

Despite the reasonableness of prior works, most current work on polarization has relied on textual content to detect sentiment and then polarization, or has been confined to specific domains 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.

3. Proposed Method

Our scope is within the context of classical collaborative filtering (CF) Recommendation algorithms that learn latent factor models, specifically Non-negative Matrix Factorization (NMF), to represent items and users based on a set of factors inferred from rating patterns.

3.1. Problem Definition

Following the polarization definition in [13], we define polarization-aware collaborative filtering as follows:

Definition 1 - 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. From definition 2, an instance from (U, I, R) can be denoted by (u, i, r) which means that user u rated item i with value r.

3.2. Polarization-aware Recommender Interactive System - (PaRIS)

The general problem of Non-negative matrix Factorization (NMF) is to decompose a non-negative data matrix R with size $m \times n$ into two positive elements matrix factors P and Q, with size $m \times k$ and $k \times n$, respectively, such that $k \ll \min(n, m)$ is a positive integer representing the rank of matrices P and Q [19]. Here, our goal is to design a recommendation system which not only recommends relevant items but also includes opposite views in case the user is interested to discover new items. A classical NMF predicts the overall rating $\hat{r_{ij}} = p_i.q_j$ by minimizing $||r_{ij} - p_i.q_j||^2$. However, the item rating alone is not able to fully consider the polarization phenomenon. Hence, in order to estimate R from (U, I) such that the system considers both relevance and polarization, we propose to minimize the objective function:

$$
\begin{aligned}
\min_{p_u,\, q_i} \quad \mathcal{J}_{PaRIS} \;&= \sum_{u \in U} \sum_{i \in I} \Big((1 - \lambda_u) \times ||r_{ui} - p_u q_i||^2 + \\
&\qquad\qquad \lambda_u \times ||r'_{ui} - p_u q_i||^2 \Big) \\
r'_{ui} &= r_{ui} - \lambda_u \times (\bar{r} + \frac{g_i}{g_{max}}) \times \Phi_i^{\lambda_u + r_{ui}} \quad \text{if } r_{ui} \geq \delta \\
r'_{ui} &= r_{ui} + \lambda_u \times (\bar{r} - \frac{g_i}{g_{max}}) \times \Phi_i^{\lambda_u + r_{ui}} \quad \text{if } r_{ui} < \delta
\end{aligned}
\tag{1}
$$

where λ_u is user *u*'s discovery factor, Φ_i is item *i*'s polarization score, computed using the Polarization Detection Classifier proposed in [13], g_i measures how extreme the different viewpoints are, δ is a threshold that indicates which ratings are considered as liked versus disliked. $g_i \in [0, 1]$ 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 use 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. g_{max} is 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 is a threshold that indicates which ratings are considered as liked versus disliked. By minimizing the objective function in (1), we estimate R from (U, I) such that the system considers both relevance and polarization, but to different degrees. The first part of the optimization objective is the classical NMF optimization criterion; while the second part is the counter-polarization component. The intuition behind the second part is to bring a user and an item closer in the latent space, if the user is interested in discovering more and the item happens to be polarized. The new incremental stochastic Gradient Descent update equations at iteration $t + 1$, after each new input r_{ui}, can be derived as shown below, where e_{ui} (e_{ui}) are the ratings' (depolarized ratings') reconstruction errors.

$$
\begin{aligned}
e_{ui} &= r_{ui} - p_u^t q_i^t \\
e'_{ui} &= r'_{ui} - p_u^t q_i^t
\end{aligned}
\tag{2}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{J}_{PaRIS}}{\partial p_u} &= 2(1 - \lambda_u) \times e_{ui} \times (-q_i) + 2\lambda_u \times e'_{ui} \times (-q_i) \\
\frac{\partial \mathcal{J}_{PaRIS}}{\partial q_i} &= 2(1 - \lambda_u) \times e_{ui} \times (-p_u) + 2\lambda_u \times e'_{ui} \times (-p_u)
\end{aligned}
\tag{3}
$$

$$
\begin{aligned}
p_u^{t+1} &= p_u^t + \eta\big(2q_i^t((1-\lambda_u)e_{ui} + \lambda_u \times e'_{ui})\big) \\
q_j^{t+1} &= q_i^t + \eta\big(2p_u^t((1-\lambda_u)e_{ui} + \lambda_u \times e'_{ui})\big)
\end{aligned} \tag{4}
$$

The steps for the PaRIS algorithm are listed in Algorithm 1.

Algorithm 1 Polarization-aware Recommender Interactive System - (PaRIS)

Input: initial user-item matrix (R_{train}), f_{PDC}(Polarization classifier function [13])
Output: final user-item matrix R

For each user $u \in U$:
Repeat While u rates unrated items:
  Update model P, Q using input data R by optimizing objective function (1)
with the parameter set of λ_u, Φ_i, g_i
  $\bar{R} \leftarrow PQ$
  Find S_u which is the set of items sorted in descending order of predicted rating
  Select the top k_t items from S and recommend them to u
  User picks an item i' randomly and gives it rating $r_{ui'}$
  $R \leftarrow R \odot r_{ui'}\ for\ i' \in S_u$
   where $\odot$ is the I/O operator, meaning that user u provides rating $r_{ui'}$ for item i'

4. Experiments

For the purpose of evaluation, we resort to a simulation scenario in this preliminary work similar to [16]. We evaluate the performance of our approach in terms of rating prediction accuracy, using the Mean Squared Error (MSE) [19] and 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 [16]. 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. We generate a rating environment with 50 users and 200 items where items are evenly divided in two opposite viewpoint sets (red items and blue items). Users are also divided into two groups based on whether they like red or blue items. In order to make the environment polarized, we assume that user $u_a \in GroupA$ likes red items more and user $u_b \in GroupB$ likes blue items more than red ones. Finally, we generated environment G with different values of polarization gaps and user discovery factors.

We start by showing some experiments that illustrate examples of how an Interactive Recommender System (IRS) 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 at a time. In each iteration, we measure MSE from the training and testing phases. We also keep track of the items that a user decided to reacted to by providing a rating.

Figure 1 shows traces from the interactive recommendation system for user $u \in GroupA$,

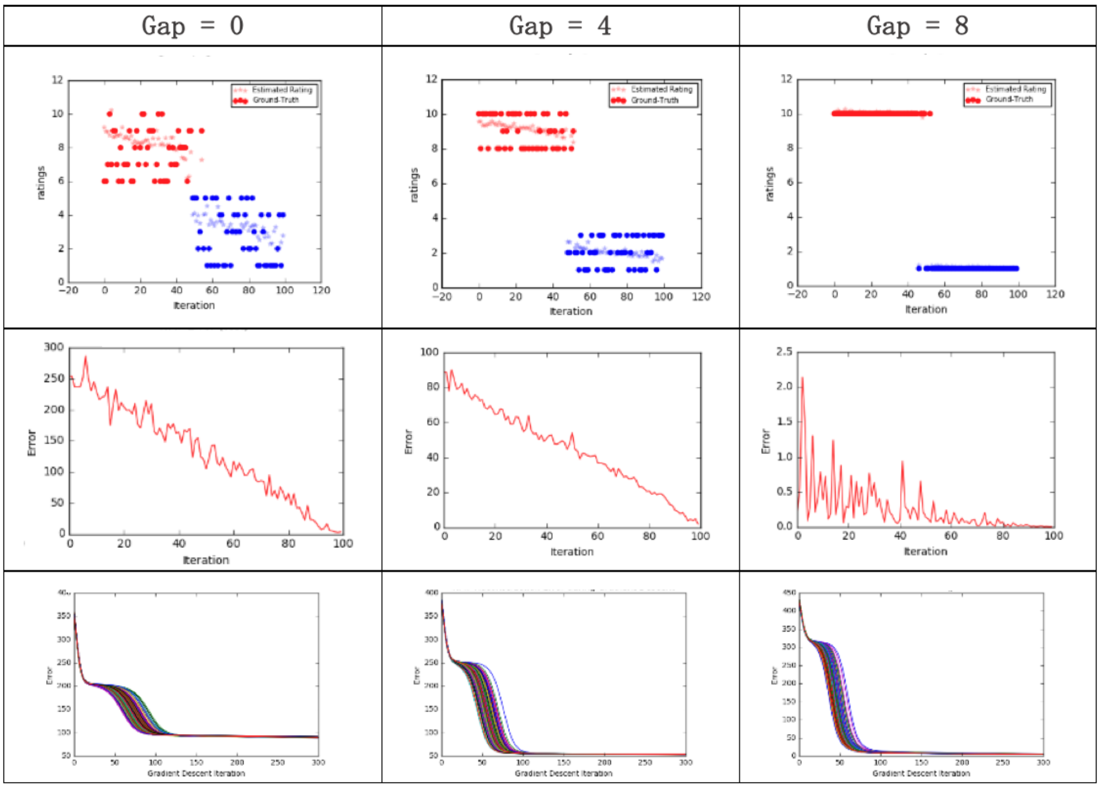

Figure 1: Traces of the Interactive Recommendation NMF-based RS in environment G with different polarization ratio and gap values, for user u_a. Polarization increases acutely with Gap. The error drops sharply in the more polarized case, making it easier for the Machine Learning model.

which means the user likes red items more than blue items. We generate environment G considering that the gap g_i is 2. Figure 1, upper row, shows that 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 them to explore any items from the opposite color/view. The second row shows the testing MSE decreases as the user provides new ratings in each iteration; hence, there are fewer unrated items for the user. Finally, the last row shows the NMF model's objective or cost function's convergence when using gradient descent optimization, where each line (or colored thread) represents the decrease of the objective function for an iteration of the interactive recommendation process.

Figure 2 shows the results of applying our proposed Polarization-aware Recommender Interactive System (PaRIS) in environment G for user u_a. As we can see, the user gets to see items from a different color/viewpoint even in a very polarized environment. The middle row shows the testing MSE error for user u_a where there are some fluctuations in the testing MSE error which is due to the modification in the main updating function of NMF. Finally, the last row shows the contrast between the objective function evolution for varying polarization rates, compared to non polarization. Furthermore the error converges to a smaller value for higher polarization. We interpret this algorithmically, by the fact that the higher the polarization in

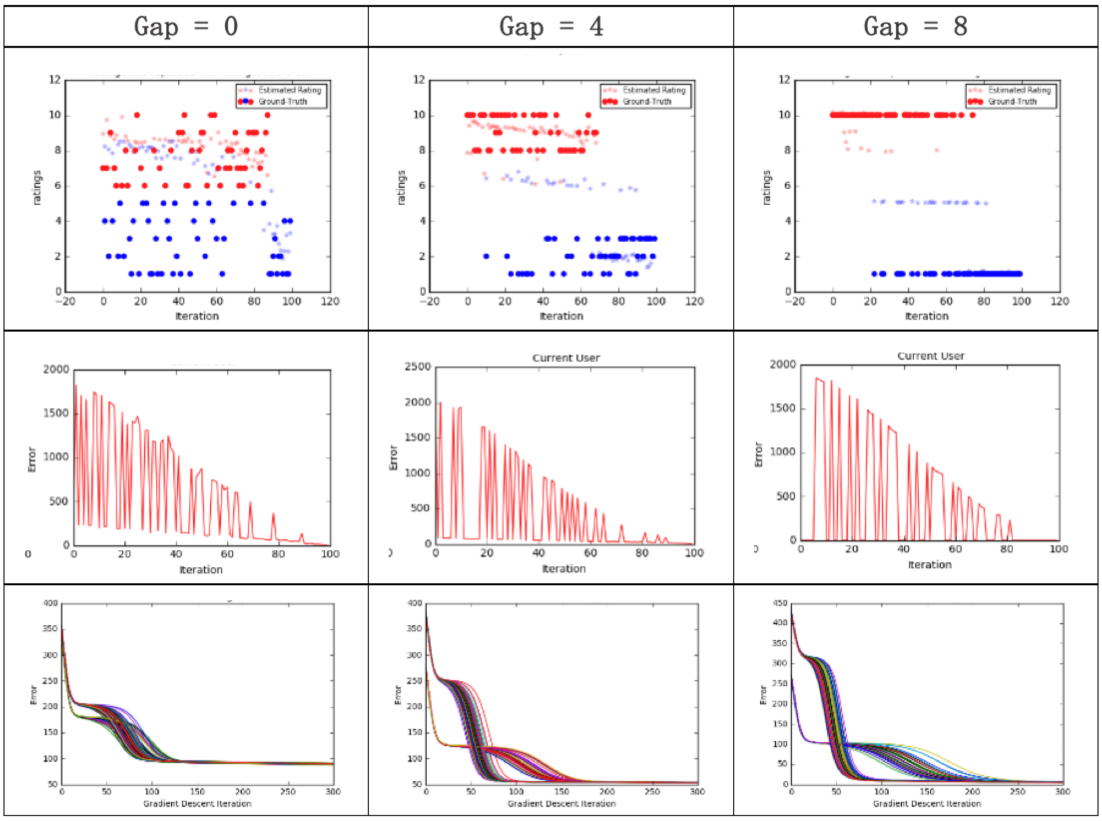

Figure 2: Traces of the Interactive Recommendation process with the polarization-aware recommender system (PaRIS) in environment G with different polarization ratio and gap values, for user u_a who had liked red items more than blue items. Polarization is reduced relative to NMF in Fig. 1 (top row). As Gap increases, the prediction error drops according to two modes: one as sharp as the unmitigated polarized case of NMF, and another decreasing more slowly like the unpolarized case at Gap = 0).

the ratings, the larger the gap (extreme likes and dislikes) between the items' ratings in the opposite viewpoints. Hence increased polarization leads to increased separation between the opposite viewpoint ratings, which, very naturally makes learning the ratings an easier task from a machine learning perspective.

We repeat the experiment with $gap = 2$ for 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 λ, $\lambda_u = c_u$ for user u and $\lambda_u = 0 \quad \forall u \in U - u$, where $c_u \in [0, 1]$, is a user defined constant. Then, we compute MSE_{test}, MSE_{train} and $OVHR$ in two ways: (a) $OVHR_u$: the ratio of number of items from the opposite view to what the user has picked from the recommendation list, (b) $OVHR_{t_k}$: the ratio of number of items recommended to the user from an opposite view.

Table 1 shows that the higher the user-defined counter-polarization tuning parameter λ, the more they will be recommended items from the opposite view, again as desired by the user. Even though the traditional NMF-based algorithm achieves good accuracy in rating prediction, it is not able to recommend any item from the opposite view. In contrast, our proposed algorithm,

Table 1
Comparison of PaRIS with the classical NMF RS in terms of accuracy and opposite view ratio ($OVHR_u$,$OVHR_{t_k}$)

			Opposite View Ratio		MSE_{Train}	MSE_{Test}
			$OVHR_u$	$OVHR_{t_k}$		
			mean, std	mean, std	mean, std	mean, std
Classic NMF			0.0 ± 0.00		22.02 ± 5.27	138.96 ± 12.55
PaRIS	Scenario (a)	$\lambda_i = 0.2$	$0.0\% \pm 0.00$	$0.0\% \pm 0.00$	57.63 ± 14.35	223.31 ± 32.14
		$\lambda_i = 0.5$	$0.0\% \pm 0.00$	$0.0\% \pm 0.00$	146.51 ± 41.73	632.21 ± 57.73
		$\lambda_i = 1.0$	$48.0\% \pm 0.17$	$32.0\% \pm 0.12$	383.80 ± 108.52	2015.85 ± 167.46
	Scenario (b)	$\lambda_i = 0.2$	$5.4\% \pm 0.073$	$4.9\% \pm 0.021$	123.92 ± 36.76	813.01 ± 36.76
		$\lambda_i = 0.5$	$6.0\% \pm 0.08$	$18.1\% \pm 0.21$	124.46 ± 37.29	299.82 ± 76.01
		$\lambda_i = 1.0$	$67.0\% \pm 0.24$	$68.0\% \pm 0.24$	361.77 ± 102.74	1883.50 ± 237.83

PaRIS, recommends 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.

5. Conclusions

We proposed a polarization-aware recommender system based on Non-negative Matrix Factorization 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. Our work is limited by the *simulation* setting in the experiments, since inducing polarization and testing different new strategies in real life has ethical risks and legal ramifications. Future work should find a way to conduct tests on real users, after mitigating the risks involved.

6. Acknowledgments

This work was supported by the National Science Foundation through grant IIS-154998.

References

[1] R. Baeza-Yates, Data and algorithmic bias in the web, in: Proceedings of the 8th ACM Conference on Web Science, ACM, 2016, pp. 1–1.

[2] J. Sanz-Cruzado, P. Castells, Enhancing structural diversity in social networks by recommending weak ties, in: Proceedings of the 12th ACM conference on recommender systems, 2018, pp. 233–241.

[3] E. Bakshy, S. Messing, L. A. Adamic, Exposure to ideologically diverse news and opinion on facebook, Science 348 (2015) 1130–1132.

[4] E. Pariser, The filter bubble: How the new personalized web is changing what we read and how we think, Penguin, 2011.

[5] Y. Ge, S. Zhao, H. Zhou, C. Pei, F. Sun, W. Ou, Y. Zhang, Understanding echo chambers in e-commerce recommender systems, in: Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval, 2020, pp. 2261–2270.

[6] U. Chitra, C. Musco, Analyzing the impact of filter bubbles on social network polarization, in: Proceedings of the 13th International Conference on Web Search and Data Mining, 2020, pp. 115–123.

[7] O. Nasraoui, P. Shafto, Human-algorithm interaction biases in the big data cycle: A markov chain iterated learning framework, CoRR abs/1608.07895 (2016). URL: http://arxiv.org/abs/1608.07895.

[8] J. Möller, D. Trilling, N. Helberger, 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 & Society 21 (2018) 959–977.

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

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

[11] J. N. Cohen, Exploring echo-systems: how algorithms shape immersive media environments., Journal of Media Literacy Education 10 (2018) 139–151.

[12] A.-A. Stoica, A. Chaintreau, Hegemony in social media and the effect of recommendations, in: Companion Proceedings of The 2019 World Wide Web Conference, 2019, pp. 575–580.

[13] M. Badami, O. Nasraoui, W. Sun, P. Shafto, 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, 2017.

[14] P. Victor, C. Cornelis, M. De Cock, A. Teredesai, A comparative analysis of trust-enhanced recommenders for controversial items., in: ICWSM, 2009.

[15] A. Antikacioglu, R. Ravi, Post processing recommender systems for diversity, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017, pp. 707–716.

[16] M. Badami, O. Nasraoui, P. Shafto, Prcp: Pre-recommendation counter-polarization., in: KDIR, 2018, pp. 280–287.

[17] M. Gao, H. J. Do, W.-T. Fu, Burst your bubble! an intelligent system for improving awareness of diverse social opinions, in: 23rd International Conference on Intelligent User Interfaces, 2018, pp. 371–383.

[18] N. Tintarev, S. Rostami, B. Smyth, Knowing the unknown: visualising consumption blindspots in recommender systems, in: Proceedings of the 33rd Annual ACM Symposium on Applied Computing, 2018, pp. 1396–1399.

[19] Y. Koren, R. Bell, C. Volinsky, Matrix factorization techniques for recommender systems, Computer 42 (2009).