Proceedings of the 4th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/55/20241422

Improvement research of Invariant Collaborative Filtering

Luyang Yu[1,6,†], Shanglin Han[2,7,†], Muheng He[3,*,†], Zekai Yang[4,8], Xinyue Hu[5,9]

[1]School of Mathematics and Science, Eastern China Normal University, Shanghai, 200241, China
[2]Art & Science, University of Miami, Miami, 33146, US
[3]School of Computing and Data Science, Xiamen University Malaysia Campus, Selangor, 43900, Malaysia
[4]McMaster Faculty of Engineering, McMaster University, Hamilton, L8S4L8, Canada
[5]School of Mathematical Sciences, Anhui University, Hefei, 230601, China

[6]10191511447@stu.ecnu.edu.cn
[7]Sxh2071@outlook.com
[*]Corresponding author email: CST2109187@xmu.edu.my
[8]yangz203@mcmaster.ca
[9]a02214082@stu.edu.cn
[†]These authors are co-first authors

Abstract. The rapid expansion of e-commerce has led to product oversaturation on online platforms. To help users find their preferred products in a more efficient and accurate way, many e-commerce platforms have introduced personalized recommendation systems. Collaborative filtering is one of the most successful techniques, while it's improvement, Invariant Collaborative Filtering (Inv-CF), aims to address the popularity bias problem of traditional CF models by capturing unbiased preferences that remain constant despite the change in popularity distributions. However, Inv-CF model still experiences some problems such as ignoring the influence of attention, causing performance less effective when analyzing the representation of implicit feedback. This paper explores the enhancement of Inv-CF, a recommendation system model designed to mitigate the influence of popularity bias. We defined the experimental framework and evaluated the performance of the improved Inv-CF on two benchmark datasets, Yahoo! R3 and COAT. And the results demonstrate significant performance gains over the original Inv-CF, highlighting the effectiveness of the proposed enhancements. In conclusion, this paper presents improvements to the Inv-CF model's loss functions, addressing issues related to popularity bias in collaborative filtering.

Keywords: Recommendation Systems, Invariant Collaborative Filtering, Popularity Bias, Attention.

1. Introduction

1.1. Background

With the rapid growth of e-commerce, the number of products on e-commerce platforms has increased dramatically and led to product oversaturation.[1] In the case of product oversaturation, users need to

© 2024 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/55/20241422

consume a lot of time to find their preferred objects from a vast number of products, so many e-commerce platforms have introduced personalized recommendation systems into their user search procedures, and the underlying logic of personalized recommendation systems is to track the past interaction between users and items, train with different data and algorithms eventually predict the user's future interactions (i.e., to provide personalized recommendations during the user's product search phase). So far there are many personalized recommendation systems, the most successful and effective of them is collaborative filtering,[2] which has two main types of recommendation algorithms: user-based and item-based collaborative filtering recommendation algorithms. The former is based on the assumption that if a person shares the same opinion with another one on a certain issue, they tend to share the same opinion on a different issue, similarly. The algorithm usually uses the nearest neighbor technique to find neighboring users, (KNN) [3] and then weights the target user's rating of the item. The latter is based on the point of view of a certain item, looking for a number of items that are similar to a certain item, and then weighting the target user's rating of the item. However, the collaborative filtering faces many challenges. CF algorithms need to be able to handle highly sparse data. Scaling with the number of users and items, CF algorithms need to make satisfactory recommendations in a short period of time and deal with other problems such as synonyms (same or similar items but with different names), precedence attacks, data noise and privacy preservation, as well as the problem of popularity bias. [4]

1.2. Overview

Recently, the proposal of a brand-new model-- Inv-CF (Invariant Collaborative Filtering) -- has caught our attention. [5] This model itself is designed to reduce the impact of popularity bias on collaborative filtering models. Popularity bias is unavoidable in real life, and most of the current debiasing strategies need to test the a priori information of the distribution to assess the degree of bias to learn further about the impact of popularity on the representation and thus, it may be able to reduce the bias. However, such strategies do not tackle the robustness of the CF model when facing different popularity distributions. [6] The core of the Inv-CF framework, on the other hand, is the representation of unbiased preferences, i.e., it is to find the invariants of preferences in the changing popular semantics. However, when studying the Inv-CF model, we found that there are improvements that can be made in the two loss functions, L_{rep}, and L_{aug}.

1.3. Major Work

1. L_{rep}:

The improvement focuses on the SoftMax loss, considering the attention within the negative samples, which indicates that though interactions do not occur, users may potentially share different opinions to different items. In that said, we mitigate that influence by computing a weight for each user-item negative sample and adding them to the loss to avoid ignoring the influence of attention factors within the negative samples.[7]

2. L_{aug}:

The original strategy focused on disentanglement of representation of preference and popularity. However, it only applied the invariant principle to the representation of preference. While we consider doing the same to the representation of popularity, ensuring that those two representations are independent and the influence to interaction caused by preference is robust.

2. Preliminary and Methodology

2.1. Research Background

Traditional CF models extract information from the implicit feedback of users for items, such as obtaining users' preferences through their historical interactions with items, to predict the future interactions between users and items in order to constitute a recommendation system. In comparison to the traditional CF model, the Inv-CF model aims at getting the users' "true preference" from the implicit feedback, i.e., a preference that is learnt based on two principles - invariance principle and

DOI: 10.54254/2755-2721/55/20241422

disentanglement principle. The preference information which is not affected by popularity shift-influenced, makes more reasonable predictions about future interactions to solve problems caused by the popularity shift. However, we believe that there is still some possibility of improving the strategy of the Inv-CF model, and without changing the principle, we propose the following simple optimization of the 2 main loss functions in order to make the model more reasonable and realistic and demonstrate the performance through data experiments.[8-12]

2.2. Definition of symbols

In this paper, the following notation is defined:

$D_{train} = \{(x, y)|x = (u, i), y = 1\}$ denotes the training set, where elements as user-item pairs in the existence of historical interactions;

$D_{test} = \{(x, y)|x = (u, i), y = 1\}$ denotes the training set, where elements serve as user-item pairs for which future interactions exist;

y denotes the existence of interactions, it is binary. y = 0 if negative samples, means no interactions, y = 1 if positive samples, means interactions exist;

u denotes the user individual;

i denotes the item individual;

$||$ denotes concatenation operation, which means two vectors are connected together.

2.3. L rep

2.3.1. Composition of $\boldsymbol{L_{rep}}$

In Inv-CF representation learning [5], since only implicit feedback exists in the original dataset, that is, whether there are known user-item interactions or not, the article adopts CF backbone model (e.g., MF [13], LightGCN [14]), to extract the representation of user, item:

$$(\boldsymbol{u}_r, \boldsymbol{i}_r) = f_r(u, i), \tag{1}$$

where $\boldsymbol{u}_r \in \mathbb{R}^d$, $\boldsymbol{i}_r \in \mathbb{R}^d$ denotes the d-dimensional preference representation of u, i. Similarly, there are:

$$(\boldsymbol{u}_p, \boldsymbol{i}_p) = f_p(u, i), \tag{2}$$

where $\boldsymbol{u}_p \in \mathbb{R}^d$, $\boldsymbol{i}_p \in \mathbb{R}^d$denotes the d-dimensional popularity representation of u, i;

In optimizing these two representations encoder, the article adopts an ERM strategy by applying Softmax loss l:

$$l(\boldsymbol{u}_r, \boldsymbol{i}_r) = -\sum_{(u,i)\in D_{train}} log \frac{\exp\left(s(\boldsymbol{u}_r, \boldsymbol{i}_r)/\tau\right)}{\sum_{i'\in N_u \cup \{i\}} \exp\left(s(\boldsymbol{u}_r, \boldsymbol{i}'_r)/\tau\right)}, \tag{3}$$

Where s is the similarity function, and here cosine similarity is used, $N_u = \{i'|(u, i') \notin D_{train}\}$ denotes the set of negative samples, and τ is the temperature hyperparameter in Softmax. Similarly, $l(\boldsymbol{u}_p, \boldsymbol{i}_p)$ is computed as a way to get minimize the empirical risk with ERM framework:

$$L_{rep} = l(\boldsymbol{u}_r, \boldsymbol{i}_r) + \alpha \cdot l(\boldsymbol{u}_p, \boldsymbol{i}_p), \tag{4}$$

2.3.2. Optimization of $\boldsymbol{L_{rep}}$

We note that in the above ERM framework of Inv-CF representation learning, Softmax loss l optimizes all negative samples of $s(\boldsymbol{u}_r, \boldsymbol{i}'_r), \boldsymbol{i}' \in N_u$ equally, however, such strategy is not entirely consistent with the actual situation, since it ignores the implicitness in users' interactions with negative samples. By the previous assumption, the negative samples are the (u, i) for which the presence of user-item interaction is not observed, while in fact, there may still be gaps among these negative samples. Inspired by Fu's work [15], combining the natural mechanism of attention with collaborative filtering, we propose, for

DOI: 10.54254/2755-2721/55/20241422

all negative samples, to build an MLP model to measure the probability of possible interactions for (u, i) that do not exist interactions with the principle of collaborative filtering algorithms, and denote such probability as w, which represents the weight of this negative sample.

Based on the consideration of realistic factors, we believe that a high weight will show that the Based on practical considerations, we think that a high weight will show that user has more preference to the item than other negative samples with lower weight, so we hope that the corresponding $s(\boldsymbol{u}_r, \boldsymbol{i}'_r)$ of the negative sample with higher weight will fall slower than those negative samples with lower weight, so we modify the original $l(\boldsymbol{u}_r, \boldsymbol{i}_r)$ as follow.

$$l_r(\boldsymbol{u}_r, \boldsymbol{i}_r) = -\sum_{(u,i)\in D_{train}} log \frac{\exp\left(s(\boldsymbol{u}_r, \boldsymbol{i}_r)/\tau\right)}{\sum_{i'\in N_u\cup\{i\}}(1-\mathrm{w_{i'}})\exp\left(s(\boldsymbol{u}_r, \boldsymbol{i}'_r)/\tau\right)}, \tag{5}$$

And based on the above understanding, when the weight is high, which means the user has relatively high preference for the item, the reason why no interaction occurs is naturally thought to be due to the discordance of popularity representation, so similarly, we modify the original $l(\boldsymbol{u}_p, \boldsymbol{i}_p)$ as follows:

$$l_p(\boldsymbol{u}_p, \boldsymbol{i}_p) = -\sum_{(u,i)\in D_{train}} log \frac{\exp\left(s(\boldsymbol{u}_r, \boldsymbol{i}_r)/\tau\right)}{\sum_{i'\in N_u\cup\{i\}}\mathrm{w_{i'}}\cdot\exp\left(s(\boldsymbol{u}_r, \boldsymbol{i}'_r)/\tau\right)}, \tag{6}$$

and from this we get the improved one:

$$L'_{rep} = l_r(\boldsymbol{u}_r, \boldsymbol{i}_r) + \alpha\cdot l_p(\boldsymbol{u}_p, \boldsymbol{i}_p), \ . \tag{7}$$

2.4. $\boldsymbol{L_{aug}}$

2.4.1. Composition of $\boldsymbol{L_{aug}}$

The original paper, which is based on the invariance principle, assumes that the relationship between preference information and interactions is stable and unaffected by changes in popularity information. Let $U_p = \{\boldsymbol{u}'_p|\forall u'\}$ denote the set of popularity representations of all users and let $I_p = \{\boldsymbol{i}'_p|\forall i'\}$ denote the set of popularity representations of all items. Let

$$\boldsymbol{u}^* = \boldsymbol{u}_r||\boldsymbol{u}'_p, \ \ \boldsymbol{i}^* = \boldsymbol{i}_r||\boldsymbol{i}'_p, \tag{8}$$

as augmented representations of (u, i), the article forces all (u, i) popularity-intervened versions to hold the consistent discriminative signals about interaction.

$$L'_{aug} = \mathbb{E}_{\boldsymbol{u}'_p\in\boldsymbol{U}_p}\boldsymbol{l}(\boldsymbol{u}^*, \boldsymbol{i}) + \mathbb{E}_{\boldsymbol{i}'_p\in\boldsymbol{I}_p}\boldsymbol{l}(\boldsymbol{u}, \boldsymbol{i}^*), \tag{9}$$

Here $\boldsymbol{l}$ is of the same form as in equation (1), with only the input term changed. This resulting process makes the preference representation much more robust for interaction when facing diverse popularity representations.

2.4.2. Improvements to $\boldsymbol{L_{aug}}$

The invariance principle presented in the article is clear and the optimisation strategy described above is very convincing, however, we believe that a simple refinement can be made to make the strategy even better, that is, to perform a complete pairwise operation, such that $U_r = \{\boldsymbol{u}'_r|\forall u'\}$ denotes the set of all user's preference representations, $I_r = \{\boldsymbol{i}'_r|\forall i'\}$ denotes the set of preference representations for all items. Let

$$\boldsymbol{u}^{**} = \boldsymbol{u}_p||\boldsymbol{u}'_r, \ \ \boldsymbol{i}^{**} = \boldsymbol{i}_p||\boldsymbol{i}'_r, \tag{10}$$

and modify the loss function as follows:

$$L'_{aug} = \mathbb{E}_{\boldsymbol{u}'_p\in\boldsymbol{U}_p}\boldsymbol{l}(\boldsymbol{u}^*, \boldsymbol{i}) + \mathbb{E}_{\boldsymbol{i}'_p\in\boldsymbol{I}_p}\boldsymbol{l}(\boldsymbol{u}, \boldsymbol{i}^*) - \mathbb{E}_{\boldsymbol{u}'_r\in\boldsymbol{U}_r}\boldsymbol{l}(\boldsymbol{u}^{**}, \boldsymbol{i}) - \mathbb{E}_{\boldsymbol{i}'_r\in\boldsymbol{I}_r}\boldsymbol{l}(\boldsymbol{u}, \boldsymbol{i}^{**}), \tag{11}$$

to reduce the impact of changes in popularity on the interaction, thus reinforcing the meaning of the invariance principle that preference shields Y from the influence of popularity.

DOI: 10.54254/2755-2721/55/20241422

3. Experiment

3.1. Experiment framework

The new and improved data is introduced on top of the original experiment, but the same dataset needs to be used to ensure comparative validity with the original Inv-CF, including the selection of feedback indicator criteria.

The main replicated experimental module is the RQ1 part which reflects the improvement of Inv-CF+ (improved) compared to the original Inv-CF.

In this part of the experiment, two high-performance CF models, MF and LightGCN, are still chosen as the backbone models for optimizations. We performed a full comparison of Inv-CF with two backbones and two baselines:

1.Inv-CF consistently significantly outperforms the state-of-the-art baselines in all metrics across all popularity distributions. We attribute the robustness of Inv-CF to extracting invariant preference information regardless of changes in the prevalence distribution. And the imp of the improved Inv-CF+ should be numerically improved over Inv.

2. The entanglement model exhibits instability when the popularity distribution changes. In contrast, thanks to the absence of implicit assumptions on the test data, CD^2AN, sDRO and our Inv-CF can greatly improve the performance of the backbone model. After the application of Inv-CF, the entanglement representation, and the high requirement of a priori information are two shortcomings of current debiasing methods. Therefore, it is reflected in the numerical improvement (and the degree of generalization of the accepted dataset) after the application of Inv-CF.

3.2. Definition of symbols

Two backbones (two high-performance collaborative filtering models):

1. MF - based on id
2. LightGCN - based on graph.

Two baselines:

1. Popularity de-biased baseline.

sam+reg - regularization based framework.

IPS-CN - reweighting based approach

CausE, MACR - Causal embedding methods

2. Popularity domain generalization baseline.

CD^2AN - the latest competing method which co-trains biased and unbiased models

sDRO - Worst-case performance improvement in a distributionally robust optimization framework.

Evaluation metrics:

A full ranking strategy is used, where. All samples except positive samples in the training set are ranked by the CF model for each user. Recall@K and Normalised Discounted Cumulative Gain (NDCG@K) - are used to evaluate the quality of the recommendations and K is set to 20 by default.

1. Recall: Recall is a commonly used evaluation metric for recommender systems in the recall phase. We often see Recall as an evaluation metric in other domains as well, which means how many of the positive samples are predicted to be true , $recall = \frac{TP}{TP+FN}$. Then in recommender systems, we can usually get the result of a recall by $recall_u = \frac{|R(u) \cap T(u)|}{|\cap T(u)|}$. Where u is the user, $R(u)$ is the set of items predicted by the model to be recommended, and $T(u)$ represents the set of recommended items in the real test set. The recall on the whole dataset can be obtained by averaging the recall for each user.
2. NDCG: This evaluation metric named Normalized Discounted cumulative gain directly translates to Normalized Discounted Cumulative Gain. Here the DCG is normalized by applying IDCG to it with the following formula: $NDCG = \frac{DCG}{IDCG}$, $IDCG = \sum_{i=1}^{|REL|} \frac{2^{rel_i}-1}{\log_2(i+1)}$. Where REL means the original recall set R is sorted according to the scores in descending order, and the sorted set is used

DOI: 10.54254/2755-2721/55/20241422

to calculate DCG to get IDCG. The standardized NDCG is a relative value, which makes it possible to compare even among different users. IDCG represents the ideal situation that all the items in the recall set are preferred by the users. So, the numerator part of IDCG is all 1.

3.3. Experiment result

Table 1. The performance comparison on Yahoo!R3 and COAT dataset.

	Yahoo!R3		**COAT**	
	MF		MF	
	Recall	NDCG	Recall	NDCG
Backbone	0.1063	0.0476	0.0741	0.0361
+ sam+reg	0.1198	0.0548	0.2303	0.1869
+ IPS-CN	0.1081	0.0487	0.17	0.1377
+ CausE	0.1252	0.0573	0.2004	0.1713
+ MACR	0.1243	0.0539	0.0798	0.0358
+ sDRO	0.139	0.0661	0.2012	0.1767
+ CD^2AN	0.1451	0.069	0.2325	0.1885
+ InvCF	**0.1409**	**0.0656**	**0.243**	**0.1963**
+ InvCF+	**0.1526**	**0.0711**	**0.2542**	**0.2033**
Imp.%	**8.31%**	**8.38%**	**4.60%**	**3.56%**

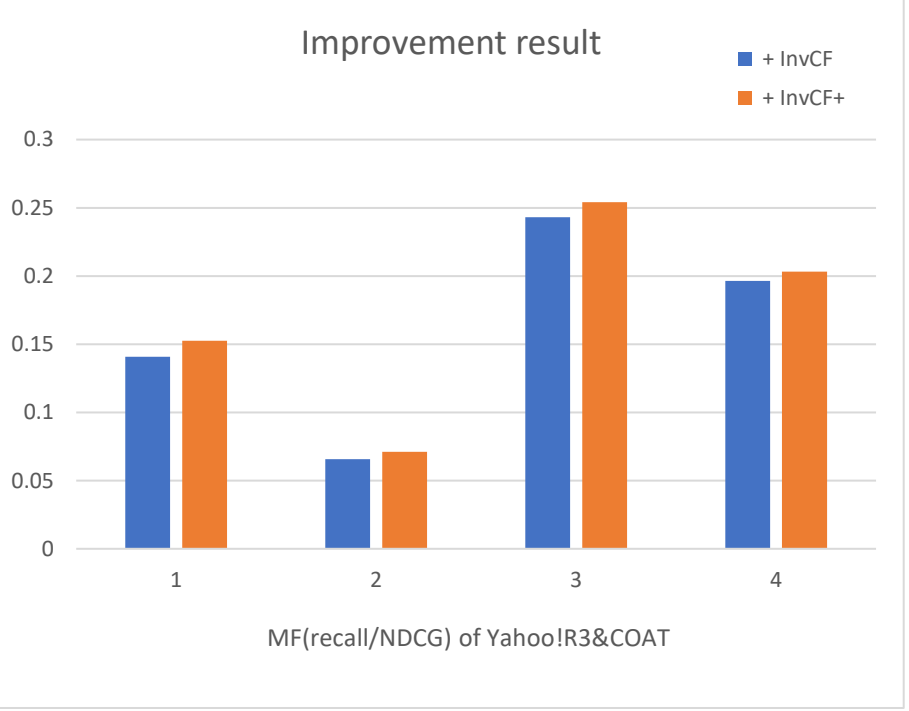

Figure 1. MF of Yahoo!R3 and COAT.

We initially carried out the improvement of the above two Loss functions for the Yahoo! R3 and COAT datasets, and its Inv-CF+ has significant performance improvement compared with the traditional Inv-CF, and there are preliminary results demonstrated for this improvement.

DOI: 10.54254/2755-2721/55/20241422

4. Conclusion

Traditional CF model nowadays is still the mainstream recommendation algorithm, which realizes its purpose by analyzing the implicit feedback of similar samples. Based on the Inv-CF model, the purpose of debias is implemented based on the invariant principle and disentanglement principle while the work of this paper focuses on the attention problem, by taking negative samples into account. Considering the reality situation, further analysis of previously undifferentiated negative samples by collaborative filtering algorithm makes the relationship between preference representations more perfect and in line with the actual psychological state of humans. Meanwhile, popularity representations are more objective, thus strengthening the disentanglement principle. Experiments show that, a better debias is implemented, which makes the model show better generalization and robustness in the face of domain adaptation. However, the performance results of the algorithm still have room for improvement. According to this paper, we believe that future work can be carried out, including examining the attention weight of negative samples by dynamic programming and strengthening the generalization of the mode for arbitrary dimensional variables, so as to further enhance the performance of CF model in the recommendation system.

Acknowledgement

Luyang Yu, Shanglin Han, and Muheng He contributed equally to this work and should be considered co-first authors.

References

[1] Kim, J. K., Cho, Y. H., Kim, W. J., Kim, J. R., & Suh, J. H. (2002). A personalized recommendation procedure for Internet shopping support. Electronic commerce research and applications, 1(3-4), 301-313

[2] Xiaoting, F. G. L. (2011). Review of collaborative filtering recommender. Library and Information Service, 55(16), 126.

[3] Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003). KNN model-based approach in classification. In On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings (pp. 986-996). Springer Berlin Heidelberg.

[4] Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. Advances in artificial intelligence, 2009.

[5] Zhang, A., Zheng, J., Wang, X., Yuan, Y., & Chua, T. S. (2023, April). Invariant Collaborative Filtering to Popularity Distribution Shift. In Proceedings of the ACM Web Conference 2023 (pp. 1240-1251).

[6] Abdollahpouri, H., Mansoury, M., Burke, R., & Mobasher, B. (2019). The unfairness of popularity bias in recommendation. arXiv preprint arXiv:1907.13286.

[7] Jang, E., Gu, S., & Poole, B. (2016). Categorical reparameterization with gumbel-softmax. arXiv preprint arXiv:1611.01144.

[8] Abdollahpouri, H., Burke, R., & Mobasher, B. (2017, August). Controlling popularity bias in learning-to-rank recommendation. In Proceedings of the eleventh ACM conference on recommender systems (pp. 42-46).

[9] Abdollahpouri, H., Burke, R., & Mobasher, B. (2019). Managing popularity bias in recommender systems with personalized re-ranking. arXiv preprint arXiv:1901.07555.

[10] Boratto, L., Fenu, G., & Marras, M. (2021). Connecting user and item perspectives in popularity debiasing for collaborative recommendation. Information Processing & Management, 58(1), 102387.

[11] Chen, Z., Wu, J., Li, C., Chen, J., Xiao, R., & Zhao, B. (2022, July). Co-training disentangled domain adaptation network for leveraging popularity bias in recommenders. In Proceedings

DOI: 10.54254/2755-2721/55/20241422

of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 60-69).

[12] Chen, J., Dong, H., Qiu, Y., He, X., Xin, X., Chen, L., ... & Yang, K. (2021, July). AutoDebias: Learning to debias for recommendation. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 21-30).

[13] Stefen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2012. BPR: Bayesian Personalized Ranking from Implicit Feedback. CoRR abs/1205.2618.

[14] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yong-Dong Zhang, and Meng Wang. 2020. LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation. In SIGIR

[15] Fu, M., Qu, H., Moges, D., & Lu, L. (2018). Attention based collaborative filtering. Neurocomputing, 311, 88-98.