Unbiasing Review Ratings with Tendency based Collaborative Filtering

Pranshi Yadav*
IIIT, Hyderabad
pranshiyadav@gmail.com

Priya Yadav*
JSSATE, Noida
priya.yadav252@gmail.com

Pegah Nokhiz
University of Utah
pnokhiz@cs.utah.edu

Vivek Gupta
University of Utah
vgupta@cs.utah.edu

Abstract

User-generated contents' score-based prediction and item recommendation has become an inseparable part of the online recommendation systems. The ratings allow people to express their opinions and may affect the market value of items and consumer confidence in e-commerce decisions. A major problem with the models designed for user review prediction is that they unknowingly neglect the rating bias occurring due to personal user bias preferences. We propose a tendency-based approach that models the user and item tendency for score prediction along with text review analysis with respect to ratings.

1 Introduction

Our society is increasingly relying on the digitized, aggregated opinions of others, which may be biased and easily manipulated while making decisions. Online reviews typically have a distribution of opinions, i.e., many are extremely positive/negative, and a few are moderate opinions.

However, numerous online reviews suffer from the rating bias problem. The opinions of individual reviewers may be affected when a person allows their preformed personal bias to affect the evaluation of another. Some users are very generous and do not rate an item with less than a 3 or 4 (on a scale of 5), thus introducing a positive bias in the scores. In contrast, some users do not go beyond a 1 or 2, thus introducing a negative bias. Hence, the reviews can affect the product's market positively or negatively, regardless of the actual product performance. The rating bias problem was earlier studied as user bias problem by Adomavicius et al. (2014); Guo and Dunson (2015); Abdollahpouri et al. (2017, 2019); Abdollahpouri (2019).

In this paper, we focus on extending a simple normalization approach by Wadbude et al. (2018), as shown in Figure 1, to mitigate user or item bias.

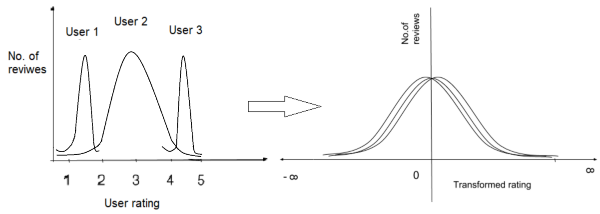

Figure 1: The user item bias problem (Wadbude et al., 2018)

Instead of using mean and standard deviation-based review rating normalization over all reviews (Wadbude et al., 2018), we perform user/item-specific review rating normalization in order to identify user/item-specific bias in ratings. That is, we will use an intuitive tendency based approach, a well-studied concept in the recommender systems literature (Sreepada et al., 2018; Cacheda et al., 2011), that estimates the bias for all user-item pairs based on user and item means/tendencies and predicts the unbiased score. Furthermore, our approach can also predict the rating of new review pairs of users and items, assuming the users or the items have a history of reviews with their corresponding ratings (i.e., no cold start setting). Therefore, we refer to Wadbude et al. (2018) for the problem introduction and Sreepada et al. (2018); Cacheda et al. (2011) for coming up with a tendency-based solution for bias mitigation and new user-item rating prediction.

The contributions in this paper are the following:

- A simple yet effective framework, inspired by Sreepada et al. (2018); Cacheda et al. (2011), based on user and item means and tendencies to underscore and mitigate the rating bias problem in reviews. Using this framework, we obtain unbiased ratings for user-item pairs.
- Furthermore, we extended the above framework to predict ratings for a new user-item pair, provided that the user or item is not novel

Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing: Student Research Workshop, pages 50–56

and has prior feedback rating (not a cold start setting). To achieve this, we come up with case-dependent reverse estimation functions.

- Through extensive experiments on publicly available datasets, we show that our approach not only helps in mitigating users' rating bias and predicting ratings for new user-item pairs, but also in detecting the exceedingly biased users/items (i.e., outliers in the review dataset) and a better alignment of predicted ratings with review text sentiments than the original scores.

In Section (1), we provided a brief introduction to the problem statement. The remaining parts of the paper are organized as follows: in Section (2) we explain our approach to predict ratings for new user-item pairs. In Section (3), we explain the experimental details. We then move on to our results and analysis in Section (4). Next, we discuss the related work in (5), followed by conclusions and future work in Section (6). We have also released the source code along with the paper.[1]

2 Rating Prediction for New User-items

We propose a user-specific statistical mapping based on user and item tendency for rating-bias removal by normalizing each review score with respect to the user and item tendencies. User tendency means whether a user tends to rate every item positively or negatively in general. Item tendency refers to whether the underlying population of users considers a specific item particularly good or bad. This is different from comparing the mean rating of the item to the global mean. The goal is to see if the item stands out among all the products rated by the user (Cacheda et al., 2011).

Let $r(c_u, p_i)$ represent the review score of user (customer) c_u for item p_i. As explained in Cacheda et al. (2011); Sreepada et al. (2018), we calculate the unbiased rating $\widehat{r_{c_u,p_i}}$ from the available ratings of user c_u and item p_i, and the predicted rating (with bias) $r(c_k, p_m)$ for a new review of user c_k for item p_m, i.e., a new user-item pair as follows:

1. Get the mean user rating $\overline{r_u}(c_u) = \frac{1}{n_p}\sum_{j=1}^{n_p} r(c_u, p_j)$ using all the reviews given by user c_u for each user. Similarly, for each item calculate the mean item rating $\overline{r_i}(p_i) = \frac{1}{n_c}\sum_{j=1}^{n_c} r(c_j, p_i)$ using the ratings given by all the users who rated that item. Here, n_p is the number of items reviewed by user c_u, and n_c is the number of reviews for item p_i.
2. Now, for every user c_u, store the user tendency $\tau_u(c_u) = \frac{1}{n_p}\sum_{i=1}^{n_p}(r(c_u, p_i) - \overline{r_i}(p_i))$ using all the reviews provided by the user c_u. Similarly, for every item p_i, we calculate the item tendency $\tau_i(p_i) = \frac{1}{n_c}\sum_{u=1}^{n_c}(r(c_u, p_i) - \overline{r_u}(c_u))$ using all the review scores given for the item p_i. Here also, n_p is the number of items reviewed by user c_u, and n_c is the number of reviews for item p_i.

   For simplicity, for rest of the paper we represent rating $r(c_u, p_i)$ as r_{ui}, user mean $\overline{r_u}(c_u)$ as $\overline{r_u}$, item mean $\overline{r_i}(p_i)$ as $\overline{r_i}$, user tendency $\tau_u(c_u)$ as τ_u, and item tendency $\tau_i(p_i)$ as τ_i. Furthermore, we will also refer to n_p as n_u, and n_c as n_i for rest of the paper.
3. To calculate the unbiased ratings $\widehat{r_{c_u,p_i}}$ for a given user-item pair (c_u, p_i), we first calculate the user and item means (i.e., $\overline{r_u}$ and $\overline{r_i}$, respectively) and user and item tendencies (i.e., τ_u and τ_i, respectively). Based on the comparison of the user mean and item mean and the sign of the user tendency and item tendency, different cases are defined for the calculation of the unbiased rating $\widehat{r_{c_u,p_i}}$ for a given user-item pair (c_u, p_i) as shown in Table 1. The β (between 0 and 1) in Table 1 is a hyper-parameter that controls the contribution of the item mean user tendency and user mean item tendency for unbiased rating calculation. For all our experiments we set β to the standard value of 0.5. For simplicity, we represent unbiased rating $\widehat{r_{c_u,p_i}}$ as $\widehat{r_{ui}}$ for the rest of the paper.

Case	Unbias Function ($\widehat{r_{ui}}$)
1. $\tau_u > 0$, $\tau_i > 0$	$\max(\overline{r_u} + \tau_i, \overline{r_i} + \tau_u)$
2. $\tau_u < 0$, $\tau_i < 0$	$\min(\overline{r_u} + \tau_i, \overline{r_i} + \tau_u)$
3. $\tau_u < 0$, $\tau_i > 0$, $\overline{r_u} < \overline{r_i}$	$\min(\max(\overline{r_u}, (\overline{r_i} + \tau_u)\beta$ $+ (\overline{r_u} + \tau_i)(1-\beta)), \overline{r_i})$
4. $\tau_u < 0$, $\tau_i > 0$, $\overline{r_u} > \overline{r_i}$	$\overline{r_i}\beta + \overline{r_u}(1-\beta)$
5. $\tau_u > 0$, $\tau_i < 0$, $\overline{r_u} > \overline{r_i}$	$\min(\max(\overline{r_i}, (\overline{r_u} + \tau_i)\beta$ $+ (\overline{r_i} + \tau_u)(1-\beta)), \overline{r_u})$
6. $\tau_u > 0$, $\tau_i < 0$, $\overline{r_u} < \overline{r_i}$	$\overline{r_u}\beta + \overline{r_i}(1-\beta)$

Table 1: Unbiasing functions (Sreepada et al., 2018)

4. Finally, we use the previous bias functions to come up with reverse estimation functions and recover the original biased rating by these

[1] https://github.com/pranshiyadav06/review-bias-normalization

functions. Reverse estimation functions are functions that restore the bias and add it to the unbiased rating for observed user/item rating predictions. We predict a review rating $r(c_k, p_m)$ (i.e., r_{km}) for the new review of user c_k for item p_m, a new user-item pair, using these reverse estimation functions. These functions are not universal and depend on user and item parameters (i.e., user/item means and tendencies). Table 2 lists the reverse functions for all the possible cases that may exist. To account for space constraints, the detailed derivations of reverse estimation functions are mentioned in appendix.[2] The ratings obtained by using the reverse estimation functions can now be compared with the given ratings (gold ratings) to evaluate the effectiveness of our tendency-based approach.

3 Experimental Details

Equations derived in Section 2, can be used to calculate the unbiased ratings for a user-item pair and predicting the ratings for a new user-item pair. We perform extensive experiments with ratings as well as the review texts associated with these ratings to evaluate our method.

3.1 Enough Rating Assumption

To empirically test this approach using real-world datasets, we need to calculate adequate invariant estimates of user/item means and tendencies from the rated data. To do so, we need to make the following assumption: The prediction example is in an online setting. Thus, the distribution means and tendencies (both users/items) will not change significantly when we predict ratings for new user-item pairs. This helps us decide the case on the estimated mean and tendency from the given labeled ratings. Similarly, the case for reverse estimation functions will be decided from prior data's means and tendencies.[3]

3.2 Dataset Details

We tested our approach on the electronics category of the SNAP Amazon e-Commerce Reviews dataset (McAuley and Leskovec, 2013). We also tested our approach on the Amazon Fine Food Reviews dataset. All the ratings in both datasets are integers from 1 to 5. In addition to review ratings, the Amazon Fine Food Reviews dataset provides the corresponding review texts and the helpfulness score of the reviews' texts. The review text could be used to obtain an alignment pattern between the differences in text reviews and their ratings. These review scores are used to get an adequate invariant estimate of each of the user/item means and tendencies from the review ratings.

Furthermore, in order to hold the assumption of an online setting, we need to have a substantial number of ratings for all users and items. To do so, we removed the items/users that have fewer ratings than a certain threshold, i.e., if the count of ratings was below 15 for Electronics and 12 for Fine Food for each user/item in both datasets. Table 3 shows the number of labeled (where the rating is provided) and unlabeled (where the rating is hidden) reviews used in the experiments after pruning. It also lists the number of users and items available in the unlabeled set. In order to evaluate the model, we have calculated the average difference in the predicted biased score and original review score for each user and item.

3.3 Experiment Procedure

Depending on the case for the mean and tendency explained in Section 2, we calculated the unbiased review score from Table 1 for the new user-item pair in the unlabeled set. To predict the review score that the user might assign to the item, we used the reverse estimation functions (Table 2, Section 2). Note that these functions will not take the minor change in the mean and tendency value that might occur due to a new user-item pair into consideration.

4 Results and Analysis

We answer the following questions through the experiments:

- Can we use a tendency-based approach to identify and remove the bias in the review ratings?
- Can we use these tendencies to predict the rating that a user may assign to an item?[4]

[2]Derivation of reverse estimation functions appendix: https://github.com/pranshiyadav06/review-bias-normalization/blob/master/Appendix.pdf

[3]One can employ a verification step after updating the means and tendencies to validate this assumption. It could be used to update the prior with the posterior distribution for modeling means and tendencies with fully Bayesian updates.

[4]The user or item is not entirely fresh and has previous reviews associated with them, i.e., no cold start setting.

Case	Sub-case	Reverse tendency estimation function
1. $\tau_u > 0, \tau_i > 0$	I. $\widehat{r_{ui}} = \overline{r_u} + \tau_i(new)$ II. $\widehat{r_{ui}} = \overline{r_i} + \tau_u(new)$	$r_{ui(new)} = (\frac{1}{n_i+n_u+1})[\widehat{r_{ui}}(n_u+1)(n_i+1) - (n_i+1)(n_u \times \tau_i) - (n_i \times n_u \times \overline{r_u})]$ $r_{ui(new)} = (\frac{1}{n_i+n_u+1})[\widehat{r_{ui}}(n_u+1)(n_i+1) - (n_u+1)(n_i \times \tau_u) - (n_i \times n_u \times \overline{r_i})]$
2. $\tau_u < 0, \tau_i < 0$	I. $\widehat{r_{ui}} = \overline{r_u} + \tau_i(new)$ II. $\widehat{r_{ui}} = \overline{r_i} + \tau_u(new)$	$r_{ui(new)} = (\frac{1}{n_i+n_u+1})[\widehat{r_{ui}}(n_u+1)(n_i+1) - (n_i+1)(n_u \times \tau_i) - (n_i \times n_u \times \overline{r_u})]$ $r_{ui(new)} = (\frac{1}{n_i+n_u+1})[\widehat{r_{ui}}(n_u+1)(n_i+1) - (n_u+1)(n_i \times \tau_u) - (n_i \times n_u \times \overline{r_i})]$
3. $\tau_u < 0$, $\tau_i > 0, \overline{r_u} < \overline{r_i}$	I. $\widehat{r_{ui}} = \overline{r_{u(final)}}$ II. $\widehat{r_{ui}} = (\overline{r_i} + \tau_u)\beta + (\overline{r_u} + \tau_i)(1-\beta)$ III. $\widehat{r_{ui}} = \overline{r_{i(final)}}$	$r_{ui(new)} = (n_i+1)\widehat{r_{ui}} - (\overline{r_u} \times n_i)$ $r_{ui(new)} = (\frac{1}{n_i+n_u+1})[(n_u+1)(n_i+1)\widehat{r_{ui}} - (n_u)(\overline{r_u} \times n_i)(1-\beta) - (n_i+1)(1-\beta)(n_u \times \tau_i) - \beta(n_i)(\overline{r_i} \times n_u) - \beta(n_u+1)(n_i \times \tau_u)]$ $r_{ui(new)} = (n_u+1)\widehat{r_{ui}} - (\overline{r_i} \times n_u)$
4. $\tau_u < 0, \tau_i > 0$, $\overline{r_u} > \overline{r_i}$	$\widehat{r_{ui}} = \overline{r_i}\beta + \overline{r_u}(1-\beta)$	$r_{ui(new)} = [\frac{1}{n_u(1-\beta)+1+\beta n_i}] \times [(n_u+1)(n_i+1)\widehat{r_{ui}} - \beta(n_i+1)(\overline{r_i} \times n_u) - (1-\beta)(n_u+1)(\overline{r_u} \times n_i)]$
5. $\tau_u > 0$, $\tau_i < 0, \overline{r_u} > \overline{r_i}$	I. $\widehat{r_{ui}} = \overline{r_i(final)}$ II. $\widehat{r_{ui}} = (\overline{r_u} + \tau_i)\beta + (\overline{r_i} + \tau_u)(1-\beta)$ III. $\widehat{r_{ui}} = \overline{r_u(final)}$	$r_{ui(new)} = (n_u+1)\widehat{r_{ui}} - (\overline{r_i} \times n_u)$ $r_{ui(new)} = (\frac{1}{n_i+n_u+1})[(n_u+1)(n_i+1)\widehat{r_{ui}} - (n_i)(\overline{r_i} \times n_u)(1-\beta) - (n_u+1)(1-\beta)(n_i \times \tau_u) - \beta(n_u)(\overline{r_u} \times n_i) - \beta(n_i+1)(n_u \times \tau_i)]$ $r_{ui(new)} = (n_i+1)\widehat{r_{ui}} - (\overline{r_u} \times n_i)$
6. $\tau_u > 0, \tau_i < 0$, $\overline{r_u} < \overline{r_i}$	$\widehat{r_{ui}} = \overline{r_u}\beta + \overline{r_i}(1-\beta)$	$r_{ui(new)} = [\frac{1}{n_i(1-\beta)+1+\beta n_u}] \times [(n_u+1)(n_i+1)\widehat{r_{ui}} - \beta(n_u+1)(\overline{r_u} \times n_i) - (1-\beta)(n_i+1)(\overline{r_i} \times n_u)]$

Table 2: Reverse estimation functions to predict review ratings (with bias) for a new user-item pair

Dataset	#Label	#Unlabel	#users	#items
Electronics	303K	75K	17486	18914
Fine Food	75K	18K	4904	3978

Table 3: Dataset statistics (after pruning)

- Can we identify anomalous ratings given the fact that a biased user can rate highly positively or negatively irrespective of the item quality?
- Can we find an alignment pattern between the difference in the reviews written for an item, the given rating, and the predicted rating?

4.1 Unbiased Ratings

We observed that for most of the user-item pairs, the unbiased review scores lie in the upper range of the rating spectrum (around 4), as shown in Figures 2 and 3. This shows that most of the people do not consistently assign poor review scores. This could also be due to the fact that more rated items are of good quality; thus, the users provide satisfactory reviews for these items.

4.2 Predicted Ratings

For the unlabeled set (hidden ratings) we calculated the predicted ratings by using the reverse estimation function (Table 2, Section 2). We then compared our predicted rating estimations with the true hidden ratings by using standard metrics such as the mean absolute error, mean squared error, and root mean squared error. Table 4 shows the error value returned on the unlabeled set for both datasets. On average, the predicted value differs from the actual value, which was assigned by the user within a range of 0.75. Therefore, our approach can accurately predict the rating of a new user-item pair.

Dataset	MAE	MSE	RMSE
Electronics	0.75	1.07	1.03
Fine Food	0.57	0.78	0.88

Table 4: Prediction rating on the unlabeled dataset

We also plotted the distribution of the difference in the predicted and the actual ratings. Figures 4 and 5 show the variation in the difference between the predicted review ratings and the actual ratings for items and users (in Electronics), respectively. The distribution is close to a normal with a mean (and peak at) zero and a standard deviation in a range of < 1.0. A similar distribution was observed for Fine Food reviews, as shown in Figures 6 and 7. Hence, we conclude that most predicted values lie within a low-margin error range of review ratings.

Table 5 represents the number of users and items with a certain average error (difference in predicted and actual review score) in various intervals of error. This shows that the majority of predicted review scores have an acceptable error of < 1.0.

4.3 Outliers

As shown in Figures 4 and 5, there are a few user-item pairs whose average error between predicted and actual ratings is substantially high. Neverthe-

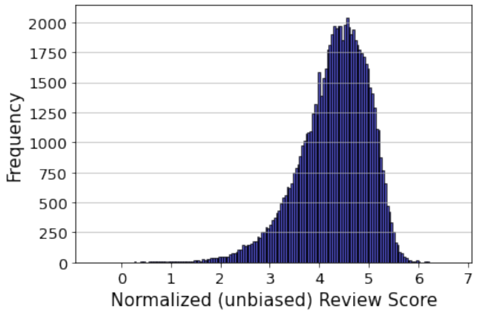

Figure 2: Unbiased rating distribution of Amazon Electronics

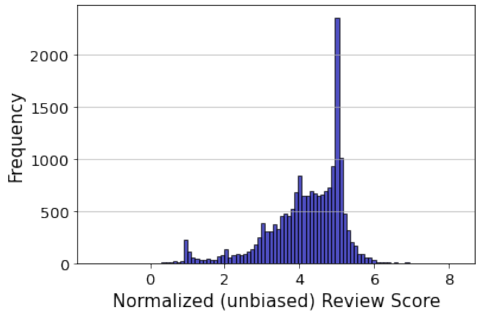

Figure 3: Unbiased rating distribution of Amazon Fine Food Reviews

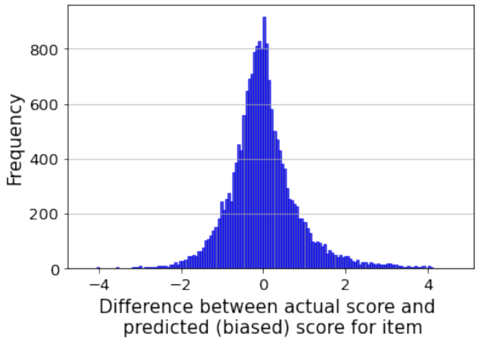

Figure 4: Average difference in predicted and Actual Review Score for items for Amazon Electronics

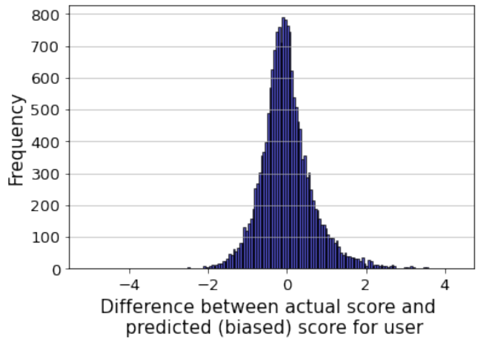

Figure 5: Average difference in predicted and Actual Review Score for users for Amazon Electronics

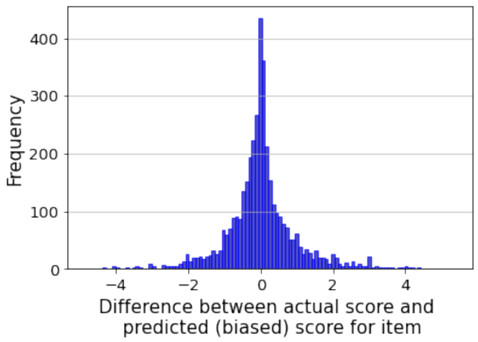

Figure 6: Average difference in predicted and Actual Review Score for items for Amazon Fine Food Review

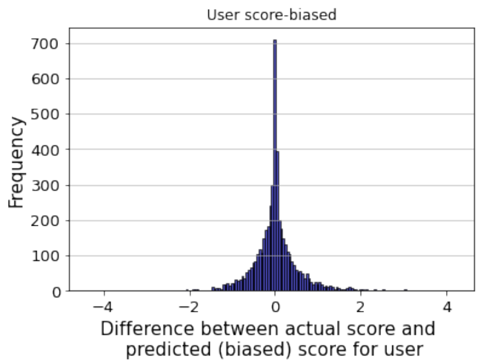

Figure 7: Average difference in predicted and Actual Review Score for users for Amazon Fine Food Review

<table>
<tr><th>Interval</th><th>#users /% of total
(Electronic | Food)</th><th>#items/% of total
(Electronics | Food)</th></tr>
<tr><td>- 0.2 – 0.2</td><td>5252/30 | 2366/48</td><td>5175/27 | 1357/34</td></tr>
<tr><td>- 0.5 – 0.5</td><td>11057/63 | 3657/75</td><td>10707/57 | 2324/58</td></tr>
<tr><td>- 1.0 – 1.0</td><td>15353/88 | 4468/91</td><td>15329/81 | 3139/80</td></tr>
<tr><td>> 2 or < 2</td><td>286/1.6 | 79/1.6</td><td>788/4 | 289/7</td></tr>
</table>

Table 5: Number of users and items for various error ranges

less, the number of such pairs is insignificant compared to the total number of pairs. Only 1.6% of users and 4% of items have such outliers (refer to Table 5). A possible reason behind these outliers is that a positively biased user can get a defective item and rate it with an extremely poor score and vice-versa. Table 6 shows some of the examples where the rating is extremely good for negatively biased (negative tendency) users and items, and extremely bad for positively biased users and items.

item / user	Bias Score	Unbias Score	Predicted Score	Diff
I1 / U1	5.0	2.68	2.75	-2.25
I2 / U2	2.0	3.68	4.06	2.06
I3 / U3	1.0	3.43	3.50	2.50
I4 / U4	1.0	5.35	5.04	4.04
I5 / U5	5.0	-0.55	-0.06	-5.06

Table 6: User-item outlier pairs' predicted ratings

4.4 Review Text Analysis

Table 7 shows that there are some user-item pairs whose review text summary is ambiguous, and the predicted score from our approach is closer to the sentiment of the text review rather than the labeled review score provided in the data. In addition, this ambiguity results in a possible range of review ratings (positive or negative) that can be assigned to the text. Therefore, the labeled rating would be the most extreme value if the user-item is biased. An explanation for this could be the presence of inherent bias in the user or item. Writing a standalone textual review mitigates the user bias compared to merely assigning a singular rating as feedback.

Summary	Orignal Score	Predict Score
doe a good job but no miracl	4.0	3.03
wonder sound high qualiti stiff cord	4.0	4.67
mix impress	3.0	3.89
super slick look use of magnet be	3.0	3.83
d solid mid level nikon dslr	4.0	4.93

Table 7: Review text analysis

Note that the review rating prediction does not take into account the sentiments expressed in the review text. That is, the reverse tendency estimation functions do not take into account the sentiment scores from the text of the review, and here, we are just using the review text to crosscheck and contrast the sentiments in the text with the original and predicted ratings.

5 Related Work

Most of the earlier approaches in rating prediction fall into three categories: a) Sentiment Analysis, b) Recommendation Systems (with/out collaborative filtering), and c) A hybrid of both. In sentiment analysis papers, the review text is used to predict the rating. This approach completely ignores the user/item meta-data, which is useful for rating bias estimation. Multiple instances of works fall in this category, but only a few are closer to our approach, e.g., Wang et al. (2016); Lei et al. (2017).

In the recommendation setting, rating predictions are treated as a matrix completion problems where the similarity between users is utilized for recommending new items. For example, if user a is similar to user b, one can assume that the rating of both users would be similar for a given item. In this setting, earlier work ignores the rating bias in users and items' tendencies. Another line of similar work, such as Musat et al. (2013); Sreepada et al. (2018); Cacheda et al. (2011) uses tendency based collaborative filtering for rating prediction. Similar to before, this line of work either ignores or does not thoroughly analyze the rating bias problem.

There are some attempts, such as Ling et al. (2014); Du et al. (2017); Jakob et al. (2009); Pero and Horvath (2013); Rao et al., which use a hybrid model for both personalized sentiment prediction and recommendation by combining information from both review and text; however, they too have not acknowledged the rating bias problem in detail. The rating bias problem was earlier studied as user bias problem by Adomavicius et al. (2014); Guo and Dunson (2015); Abdollahpouri et al. (2017, 2019); Abdollahpouri (2019). However, the most relevant work to our approach is Wadbude et al. (2018), which acknowledges the rating bias problem using simple mean/standard deviation-based normalization schemes. Although, in our paper, we use simple yet effective tendency based collaborative filtering instead of a standard normalization.

6 Conclusion

In this paper, a tendency-based approach that models the user and item tendency for rating prediction is proposed and tested on standard review datasets. To do so, we use the existing established work on tendency-based collaborative filtering for obtaining reverse estimation functions. Our method can successfully mitigate user rating bias and can also help in detecting outliers/anomalies in ratings and reviews' texts. We will further extend this model by analyzing tendencies and score predictions based on the users' demographics and items' categories.

Currently, our model cannot handle the cold start problem and require a sufficient history of earlier ratings for users/products. Thus, our model cannot predict ratings for completely new users and items. We intend to examine this problem in the future extensions of the model. Moreover, despite the bias in the text and score rating, we noted that bias is more inherent while rating a product. As shown in Table 7, we see that text can be ambiguous or neutral while the review score has a substantial amount of bias in it. We plan to extend the proposed model to take into account the positive and negative text-based sentiments along with the scores for a more accurate bias modeling. Thus, using the review text sentiments is a possible future direction. One could also use the sentiments as a means to compare with techniques that detect bias from the text.

References

Himan Abdollahpouri. 2019. Popularity bias in ranking and recommendation. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pages 529–530. ACM.

Himan Abdollahpouri, Robin Burke, and Bamshad Mobasher. 2017. Controlling popularity bias in learning-to-rank recommendation. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*, pages 42–46. ACM.

Himan Abdollahpouri, Robin Burke, and Bamshad Mobasher. 2019. Managing popularity bias in recommender systems with personalized re-ranking. *arXiv preprint arXiv:1901.07555*.

Gediminas Adomavicius, Jesse C Bockstedt, Curley Shawn, and Jingjing Zhang. 2014. De-biasing user preference ratings in recommender systems. In *Joint Workshop on Interfaces and Human Decision Making for Recommender Systems, IntRS 2014, Co-located with ACM Conference on Recommender Systems, RecSys 2014*, pages 2–9. CEUR-WS.

Fidel Cacheda, Víctor Carneiro, Diego Fernández, and Vreixo Formoso. 2011. Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems. *ACM Transactions on the Web (TWEB)*, 5(1):1–33.

Yali Du, Chang Xu, and Dacheng Tao. 2017. Privileged matrix factorization for collaborative filtering. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)*, pages 1610–1616. IJCAI.

Fangjian Guo and David B Dunson. 2015. Uncovering systematic bias in ratings across categories: A bayesian approach. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 317–320.

Niklas Jakob, Stefan Hagen Weber, Mark Christoph Muller, and Iryna Gurevych. 2009. Beyond the stars: exploiting free-text user reviews to improve the accuracy of movie recommendations. In *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, pages 57–64. CIKM.

Zheng Lei, Vahid R. Noroozi, , and Philip S. Yu. 2017. Joint deep modeling of users and items using reviews for recommendation. In *Proceedings of the tenth ACM International Conference on Web Search and Data*, pages 425–434. ACM.

Guang Ling, Michael R. Lyu, , and Irwin King. 2014. Ratings meet reviews, a combined approach to recommend. In *Proceedings of the 8th ACM Conference on Recommender systems*, pages 105–112. ACM.

Julian McAuley and Jure Leskovec. 2013. Hidden factors and hidden topics: understanding rating dimensions with review text. In *Proceedings of the 7th ACM conference on Recommender systems*, pages 165–172. ACM.

Claudiu-Cristian Musat, Yizhong Liang, and Boi Faltings. 2013. Recommendation using textual opinions. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, pages 2684–2690. IJCAI.

Stefan Pero and Tomas Horvath. 2013. Opinion-driven matrix factorization for rating prediction. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 1–13. UMAP.

K Yogeswara Rao, GSN Murthy, and S Adinarayana. Product recommendation system from users reviews using sentiment analysis. *International Journal of Computer Applications*, 975:8887.

Rama Syamala Sreepada, Bidyut Kr Patra, Avijit Chakrabarty, and Satyadev Chandak. 2018. Revisiting tendency based collaborative filtering for personalized recommendations. In *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data*, pages 230–239. ACM.

Rahul Wadbude, Vivek Gupta, Dheeraj Mekala, and Harish Karnick. 2018. User bias removal in review score prediction. In *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data*, pages 175–179. ACM.

Bingkun Wang, Yongfeng Huang, and Xing Li. 2016. Combining review text content and reviewer-item rating matrix to predict review rating. In *Computational Intelligence and Neuroscience*. Hindawi.