Normalization of Bias Score for Review Prediction

Pranshi Yadav , Priya Yadav*
JSSATE, Noida
{pranshiyadav,priya.yadav252}@gmail.com

Pegah Nokhiz, Vivek Gupta
School of Computing, University of Utah
{pnokhiz,vgupta}@cs.utah.edu

ABSTRACT

User-generated content's score based prediction and product 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 completely neglect the user bias occurring due to personal user bias preferences. We propose a tendency-based approach that model the user tendency along with text review analysis for score prediction.

ACM Reference format:
Pranshi Yadav , Priya Yadav and Pegah Nokhiz, Vivek Gupta. 2020. Normalization of Bias Score for Review Prediction. In *Proceedings of The ACM India Joint International Conference on Data Science & Management of Data, Hyderabad, India, January 5–7, 2020 (CoDS-COMAD '20),* 1 pages.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

User Bias Problem: In decision making, our society is increasingly relying on the digitized, aggregated opinions of others which may be biased and easily manipulated. Reviews typically have a distribution of opinions with many extreme positive and/or negative reviews, and few moderate opinions. 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 less than 3 or 4 (on a scale of 5), thus introducing a positive bias in the scores. On the other side, some users do not go beyond 1 or 2 (on a scale of 5), thus introducing a negative bias. Hence, the review can affect the market positively or negatively regardless of the actual performance.

2 RELATED WORK

Past work that perform user-bias modelling only focus on collaborative filtering which is complex. On the other side, for a global classifier, the average rating of the product is considered to be unbiased. When the users are biased, there is a high probability that the average product rating is also biased. Other works which deal with bias in review score are mostly concerned with the user [3]. In our approach, we use an intuitive tendency based model that estimates the user bias for all users based on user and product tendency. Instead of normalizing over all reviews, like earlier work does, we do user specific normalization in order to identify user-specific bias. Finally, we learn a global regressor to predict unknown scores given corresponding reviews.

*First two authors are undergrad students at JSSATE, Noida and contributed equally

CoDS-COMAD '20, January 5–7, 2020, Hyderabad, India
2020. ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

3 USER-BIAS REMOVAL

We propose to develop a user specific statistical mapping based on user and product tendency for user-bias removal, by normalizing each review score with respect to the user and product tendency. Let $R(u_i, p_j)$ represent the review score of user u_i for product p_j. We calculate the normalized score $NR(u_i, p_j)$ for training, and predict score $PR(u_k, p_m)$ for new review of user u_k for product p_m during prediction as follows:

(1) For each user store the mean $R_\mu(u_i) = \frac{1}{N_{u_i}} \sum_{j=1}^{N_{u_i}} R(u_i, p_j)$ of all scores given by user u_i. Similarly, for each product store the mean $R_\mu(p_i) = \frac{1}{N_{p_i}} \sum_{j=1}^{N_{p_i}} R(u_j, p_i)$ of the scores given by all the users. Here, N_{u_i} is the number of products reviewed by user u_i, N_{p_i} is the number of reviews for product p_i.
(2) Now, for every user u_i, store the user tendency $\mathcal{T}_u(u_i) = \frac{1}{N_{u_i}} \sum_{j=1}^{N_{u_i}} R(u_i, p_j) - R_\mu(u_i)$ of all the review scores given by the user u_i. Similarly, for every product p_j, store the product tendency $\mathcal{T}_p(p_i) = \frac{1}{N_{p_j}} \sum_{i=1}^{N_{p_i}} R(u_j, p_i) - R_\mu(p_i)$ of all the review scores given for the product p_j.
(3) The above-calculated tendencies can be positive or negative, and based on their values, different cases [2] are defined for the calculation of normalized score $NR(u_i, p_j)$ for each user-product pair. For details refer to the footnote link. [1]
(4) During prediction, the regressor predict a normalized review rating $PNR(u_k, p_m)$ for new review of user u_k for product p_m. We recover the original biased score by the reverse tendency based estimation. For details refer to the footnote link. [2]

4 CURRENT WORK

Currently, we are testing our approach on multiple categories of the SNAP Amazon e-Commerce Reviews dataset [1]. The dataset consists of the review score and review text along with the helpfulness parameter of the text of 1.4 million reviews, which can be used for score and text analysis of the reviews. We will further extend this model by analysing tendencies and score predictions on the basis of the demographics of users and the categories of the products.

REFERENCES

[1] 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*. ACM, 165–172.
[2] 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*. ACM, 230–239.
[3] 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*. ACM, 175–179.

[1]Link to normalise score estimation: https://bit.ly/2kMjU67
[2]Link to reverse tendency bias score estimation: https://bit.ly/2kMjU67