Review Ranking to Support Selection of Recommended Items

Short Paper

Taketoshi Ushiama
Faculty of Design
Kyushu University
Fukuoka, Japan
email: ushiama@design.kyushu-u.ac.jp

Daichi Minami
Graduate School of Design
Kyushu University
Fukuoka, Japan
email: minami@kyudai.jp

***Abstract*—This paper proposes a novel approach to aid product selection in e-commerce through the effective ranking of online reviews. Often, users find it challenging to identify the most valuable information amidst a sea of reviews. Our approach addresses this by ranking reviews based on the user's empathy towards reviewers. By taking user feedback on reviews of known products, we estimate the level of empathy towards the reviewer, subsequently ranking reviews of unknown items accordingly. This enables users to easily pinpoint the most relevant reviews amidst the multitude of information. Our evaluation experiments have revealed this new approach to be superior to traditional comparative methods.**

Keywords - online reviews; recommendations; rankings; natural language processing; machine learning.

I. Introduction

In recent years, with the spread of the Internet, various web services such as E-commerce sites and social media have appeared. It has become common to purchase products and research product information online. In addition, the number of items present on the Internet is increasing, and such "information overload" has become an important issue in recent years [1], [2]. Under these circumstances, it becomes difficult for users to discover items on the Internet that match their preferences. To solve this problem, many web services often provide information recommendation functions.

In recent years, social media-style Web sites (online review sites) that collect reviews about items in specific fields have attracted many users. Many e-commerce sites also offer the ability for users to post reviews on items, and many reviews on each item have been posted. Reviews play an important role in users' selection of items. However, when many reviews about the same item are included, it is impractical to browse through all reviews [3]. It has been reported that 80% of users read only a maximum of 10 reviews when purchasing an item on online review sites such as Amazon. Therefore, functions that rank reviews are essential to assist users in merchandising recommended items [1].

Some reviews are helpful to the user, while others are not. Therefore, e-commerce sites provide a mechanism for rating reviews and ranking highly rated reviews at the top. However, there exist cases where reviews that are valuable to one user are not valuable to another. Since the existing review ranking method is not personalized, reviews that are not valuable to a user may appear at the top of the list. Therefore, the review ranking mechanism is expected to reflect the values and preferences of users.

This paper proposes a method for recommending online reviews about a target item based on the user's empathy. We assume that reviews that are useful for item evaluation are reviews by reviewers who have a high degree of empathy with the target user. The proposed method predicts the target user's confidence level about a reviewer based on the reviews the user has rated in the past. The proposed method then considers the opinions of reviewers with whom the target user can empathize based on the user's level of trust in the reviewer to be of high value, estimates the value of the review to the target user, and recommends reviews based on that value.

Section II describes related works. Section III explains the trust-based collaborative filtering for reviewers that forms the background for this study. Section IV introduces a method for ranking item reviews based on empathy. Section V shows the experimental results for evaluating the effectiveness of the proposed method. Section VI describes the summary and future work.

II. Related Work

A. Information Recommendation using Reviews

In recent years, many studies on information recommendation using reviews have been conducted [4]–[15]. There are three main directions of recommendation using reviews [16].

- Tries to solve the data sparseness problem by extracting user preferences information.
- Tries to solve the cold-start problem of the inability to make high-performance recommendations when user evaluations are not sufficiently collected.
- Tries to derive useful information for a recommendation other than evaluation values, such as user context and latent preference factors, from the reviews.

Several methods have been proposed for extracting user preferences and recommending information using reviews. Hayashi et al. [17] extract interest words representing user preferences and their polarity from user-written reviews and recommend movies with reviews that contain interest words and have matching polarity. In the above information recommendation method using reviews, recommendations are made based on the degree to which the recommended items match the preferences of the target user. They do not support the selection of items recommended to the user by the recommendation system. The purpose of this research is to support the act of sorting recommended items by the user.

The approach of acquiring user preferences by having users directly input reviews for items they have selected in the past is not realistic because it places a heavy burden on the user. Therefore, the proposed method adopts an approach to estimate user preferences indirectly by using feedback from user reviews.

B. Review Ranking Methodology

Amazon has a voting button for whether a review is helpful. Users vote for reviews, and the quality of the reviews is determined based on the results. However, only about 10% of all reviews on Amazon are rated [18]. In addition, older reviews have many votes and appear at the top of the list, while newer reviews that do not yet have votes are considered useless [19].

Reviews with many positive votes are more likely to attract more positive votes. To solve these problems, studies have been conducted to estimate the quality score of reviews [20].

Some studies have been conducted on the reliability of textual information such as reviews [21]–[23]. The main purpose of these studies is to address the problem of malicious contributors who post unfair reviews, such as spam, and to automatically determine whether a post is a spam or not or whether a post is fake or not. Thus, most of the conventional studies on ranking and filtering of reviews are concerned with the objective value of reviews, and not much discussion has been given to personal preferences for reviews, such as “whether they are useful to the target user” or “whether they are favorable to the target user”. This paper differs from the above studies in that it focuses on the value of reviews to each individual.

III. Collaborative Filtering Based on the Trust in Reviewers

In general, collaborative filtering methods can be classified into user-based collaborative filtering and item-based collaborative filtering. User-based collaborative filtering calculates user similarity from a user-item evaluation matrix based on the assumption that “users who select the same item have similar preferences in item selection”. For example, “Dokusho Meter” [24], which is one of the leading online review sites in Japan, allows users to register the books they have read, so the similarity between users can be calculated based on the books they have read. It then recommends books that are likely to be of interest to the target user based on their similarity.

However, when looking at or thinking about various things, including books, people evaluate them from their own standpoints. How people perceive a subject often differs depending on their individual sensitivities. In online review sites, how users perceive an item is expressed in their reviews. Even if the reviews are about the same item, there are often differences in the topic of each review. For example, in the case of a review of a comic book, some reviewers evaluate the drawings, while others evaluate the story. This suggests that not all reviewers are sympathetic to the user, but sometimes some reviewers are not sympathetic to the user. Therefore, even among users who have selected the same item, there may be cases where their tastes are not similar.

Therefore, when recommending book information using collaborative filtering on online review sites for books, it is possible to improve the recommendation accuracy by changing the weights of reviewers and calculating the recommended books based on the target user’s preferences for reviewers [25].

Figure 1 shows a diagram of the recommendation system using trust information. First, on a page related to a book selected by the target user, reviews posted for the book are displayed. The target user reads the reviews and enters a rating for each reviewer as to whether or not he or she supports the reviewer. By calculating the similarity between the features of the rated reviewers and those of the reviewers who have not yet been rated, we estimate the target user’s confidence in the unknown reviewers. Reviewer features are vectorized from submitted documents using the pre-trained Doc2Vec [26]. Then, the recommended books are determined by using the confidence level of each reviewer as the reviewer’s weight.

The predicted evaluation value of book B_2 for user u when book B_1 is selected is calculated by the following equation.

$$\mathrm{Pred}(u, b_1, b_2) = \sum_{r \in \mathbf{R}(b_1)} \big(\mathrm{trust}(u, r) \times \mathrm{read}(r, b_2)\big) \quad (1)$$

where $\mathbf{R}(b)$ represents the set of reviewers who posted reviews on book b and $\mathrm{trust}(u, r)$ represents the trust level of reviewer r for target user u. Also, $\mathrm{read}(r, b)$ is determined by whether or not reviewer r has read book b.

$$\mathrm{read}(r, b) = \begin{cases} 1 & (r \text{ have read } b.) \\ 0 & (r \text{ haven't read } b.) \end{cases} \quad (2)$$

The sum of the trust of the reviewers who read book b_2 among the set of reviewers who read book b_1 becomes the predicted rating of book b_2 for user u, and thus the recommendation by collaborative filtering considering the user’s trust level.

IV. Proposed Method

A. Evaluation of Reviews for Known Items

The proposed method assumes that users rate reviews. Figure 2 shows an example of an online review site’s input interface for rating reviews. The user gives a rating for a review of a known item on this interface. In this figure, the two icons at the bottom right of the review are buttons for entering ratings. The “Sympathetic” icon indicates a positive rating, and the “Not sympathetic” icon indicates a negative rating. If the user’s evaluation of the review is neither “Like” nor “Dislike” the user does not click either icon. The system registers the review ratings entered by the user as user profile information.

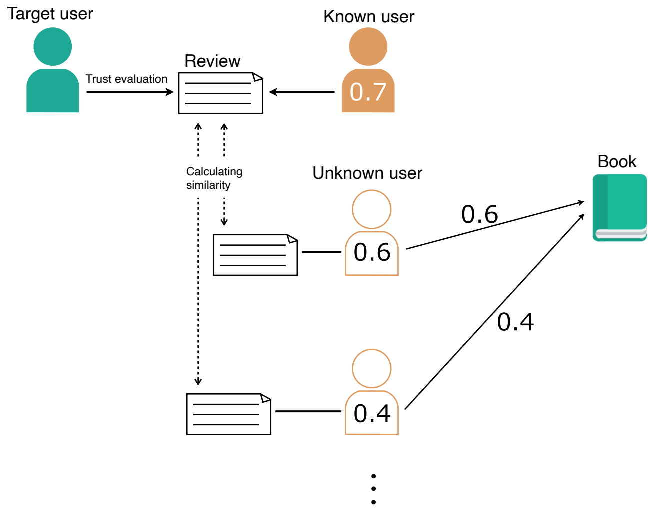

Fig. 1. Recommendation system using trust

Fig. 2. Example of a review rating on the interface

B. Value Estimation for Evaluated Reviewers

This section describes a method for estimating the value of a user's rating to a reviewer based on the rating information described in Section IV-A.

In the proposed method, users evaluate the reviews posted by reviewers, and based on the evaluations, the reviewers' trust in the users is estimated and used for recommending items. Typically, reviewers post reviews for a single item or multiple items. Each review posted by the same reviewer is expected to reflect the reviewer's characteristics. However, as reviewers post reviews for various items, they may post reviews with content they would not usually post. For example, if a review that a user gave a negative rating was of a type that the author of that review does not often post, it does not mean that other reviews written by that reviewer are less valuable to the user.

We propose a method for estimating the trust of a reviewer that considers the certainty of how reviewer-like a given rated text is in its utterances when calculating the value of a reviewer. The confidence of a reviewer is a value that is higher when a given review is a reviewer-like text and lower when it is not reviewer-like. It is used to determine how much the reviewer's confidence value should reflect the reviewer's evaluation of the review itself when computing the confidence value for the reviewer. Figure 3 is a conceptual diagram for estimating the value of a reviewer whose review is directly rated. The trust level $\mathrm{trust}(u, r)$ for a reviewer r who posted a review rated by a user u is defined by the following equation based on the rating.

$$\mathrm{trust}(u, r) = \sum_{dinD_r(pos)} \mathrm{conf}(d, r) - \sum_{dinD_r(neg)} \mathrm{conf}(d, r) \tag{3}$$

This formula estimates the user's trust of a reviewer by subtracting the sum of the confidence levels of the negatively rated document sets from the sum of the confidence levels of the document sets that received positive ratings from the user. In this equation, $D_r(pos)$ represents the set of documents posted by reviewer r that received positive ratings from users, $D_r(neg)$ represents the set of documents posted by reviewer r that received negative ratings from users, and d represents a single document. This study defines $\mathrm{conf}(d, r)$ as the confidence of document d in reviewer r. The details of the confidence level are described in Section IV-D.

C. Value Estimation for Unevaluated Reviewers

This section describes a method for estimating the trust of reviewers who have not yet been evaluated by the user among the reviewers of the item selected by the user in Section IV-A. The method calculates the cosine similarity between the feature vectors of the reviewers that have been rated by the user and the feature vectors of the reviewers that have not yet been rated. This value affects the trust of reviewers who have not yet been evaluated. The method is based on the idea that the value of a reviewer who is similar to a reviewer with high trust is high, while the value of a reviewer who is similar to a reviewer with low trust is low. Figure 4 is a conceptual diagram for estimating the value of an unevaluated reviewer. The trust $\mathrm{trust}(u, r)$ of any reviewer r for a user u is calculated using the following formula.

$$\mathrm{trust}(u, r) = \left(\sum_{d \in D(pos)} \mathrm{sim}(d, d_r) - \sum_{d \in D(neg)} \mathrm{sim}(d, d_r) \right) \times \mathrm{conf}(d_r, r) \tag{4}$$

In this formula, we first obtain the value obtained by subtracting the sum of the similarity between the reviews that user u has rated negatively and the review d_r of unrated reviewer r from the sum of the similarity between the reviews that user u has rated positively and review d_r. Then, by multiplying the obtained value by the confidence of the review, we estimate user u's confidence in reviewer r. Where $\mathbf{D}(pos)$ is the set of reviews that received positive ratings from users, and $\mathbf{D(neg)}$ is the set of reviews that received negative evaluations from the users. $\mathrm{sim}(d, d_r)$ represents the similarity between a review d and a review d_r of an unrated reviewer r, and is obtained by computing the Cosine similarity between $\mathbf{d}$ and $\mathbf{d_r}$, which are vectorized by Doc2Vec. The similarity is obtained by calculating the cosine similarity between $\mathbf{d}$ and $\mathbf{d_r}$ vectorized by Doc2Vec.

D. Computing the Confidence of a Review Using an Author-Estimation Model

This section describes a method for computing the confidence level of a review using an author estimation model for reviews based on Deep Learning. Specifically, a neural network with two hidden layers is used to train the model using a document vector that represents the semantic and lexical information of a single review using Doc2Vec and character-based unigrams as input data, and ID information associated with the user as the correct answer label. The output obtained by inputting arbitrary reviews to the trained model is the probability for each reviewer. This probability is higher when the review is a document that is typical of the target reviewer and lower when the review is not typical of the reviewer.

E. Ranking of Reviews for a Recommended Item

It is thought that the reviews that are useful in the item selection process are the reviews of reviewers with whom the user can empathize. This section proposes a method to support efficient item selection by estimating the reviews that users can relate to and displaying them at the top of the list. Figure 5 shows a conceptual diagram of the proposed method. Based on the feedback of ratings on reviews of known items A, we estimate the value of reviews of recommended items B using the confidence values for reviewers obtained in Section IV-B. Based on the assumption that reviewers who are similar to reviewers with high trust values can also be trusted, the recommendation score of review d for user u is calculated by the following formula.

$$\mathrm{score}(u, d) = \frac{\sum_{r \in RVR(PR,NR)} \mathrm{trust}(u, r) \times \mathrm{sim}(r, r_d)}{\sum_{r \in RVR(PR,NR)} \mathrm{sim}(r, r_d)} \tag{5}$$

where PR represents the set of reviews posted by user u that received positive ratings from users, and NR represents the set of reviews posted by user u that received negative ratings from users. The R_D represents the reviewer who posted review D. By multiplying the similarity between reviewer r and reviewer r_d by the confidence value of reviewer r and calculating the weighted average, we can compute the recommendation score of review d for user u. The recommendation score of d is calculated by multiplying the similarity between reviewers r and r_d and calculating the weighted average. The similarity $\mathrm{sim}(r, r_d)$ between reviewers is calculated using Doc2Vec [26], which can acquire a distributed representation of sentences. Specifically, all reviews submitted by reviewers are concatenated into a single document and converted into a vector by Doc2Vec, and the similarity between vectors is calculated by cosine similarity. The reviewer set RVR is determined by the following formula:

$$\mathrm{RVR}(PR, NR) = \bigcup_{d \in PR \cup NR} \mathrm{reviewer}(\mathrm{item}(d)) \tag{6}$$

Where $\mathrm{item}(d)$ denotes the item set to which d reviews are posted and $\mathrm{reviewer}(\mathrm{item}(d))$ denotes the set of reviewers who post reviews on the item set to which d reviews are posted.

V. Evaluation

A. Experimental Setup

Experiments were conducted to evaluate the effectiveness of the proposed method. The dataset used for the experiments

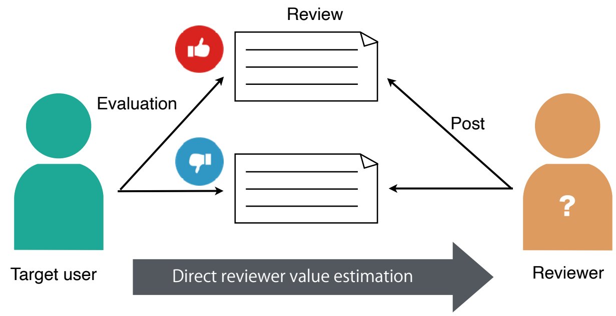

Fig. 3. Value estimation for Evaluated Reviewers

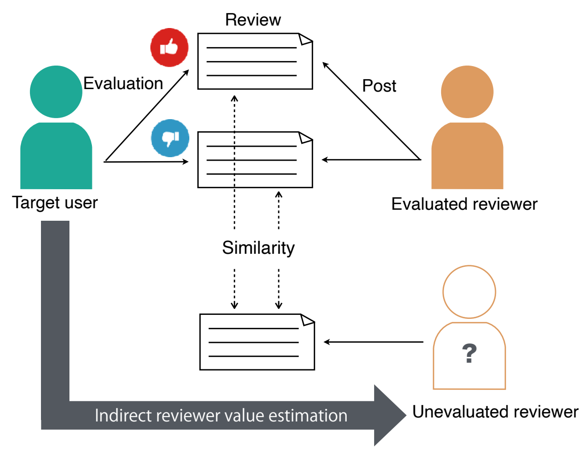

Fig. 4. Value estimation for unevaluated reviewers

was obtained by crawling from one of the famous Japanese online book review sites "Dokusho meter". MeCab[1] was used for Japanese analysis, and mecab-ipadic-neologd[2] was used for the dictionary. The number of training epochs for the author estimation model was set to 30.

The participants were asked to select two books from among those they had recently read and to input their evaluation feedback for the reviews of the two books. The participants were asked to rate the reviews of two books on a three-point scale of "agree," "don't know," and "don't agree," based on the question "do you agree with this review?" The evaluation data for one book review was used as training data to predict the recommendation score for the other book review.

The experiment participants were six men and six women in their twenties. The average of the recommendation results for 12 samples crossed between the training data and the validation data was calculated. In addition, a validation test was conducted on the accuracy of the proposed author estimation method on the reviewers of the books selected by the participants.

To evaluate the effectiveness of the proposed method, we compared the results with those of three different methods. They are random sampling, vote ranking, and Support Vector Regression (SVR). In the vote ranking, we compared the top 10 reviews with the highest number of votes with the top 10 reviews using the proposed method. In SVR, the explanatory variables for the regression analysis were the Term Frequency–Inverse Document Frequency (TF-IDF) vector of words, the percentage of each part of speech in the reviews, the total number of words, and the number of word types.

[1]https://taku910.github.io/mecab/
[2]https://github.com/neologd/mecab-ipadic-neologd

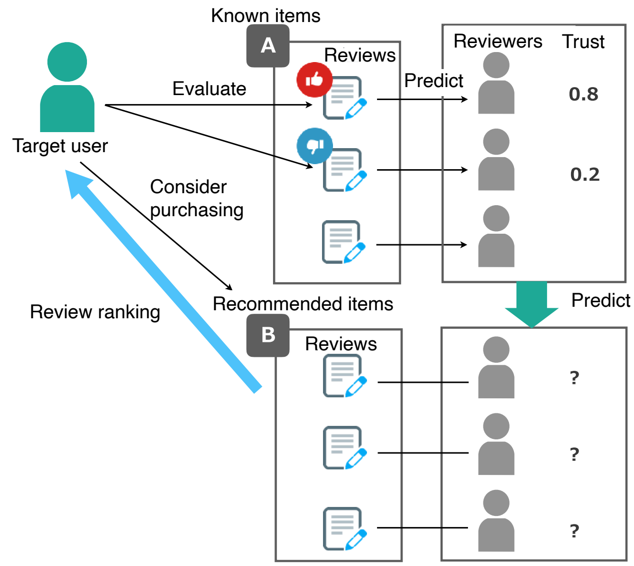

Fig. 5. Conceptual diagram of the proposed method

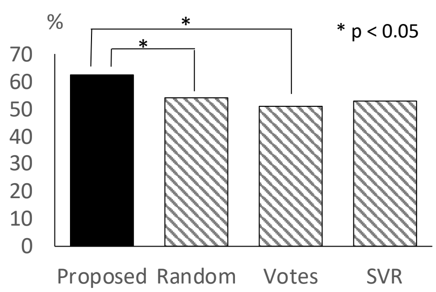

Fig. 6. Percentage of ”sympathetic” reviews in top 10 ranked.

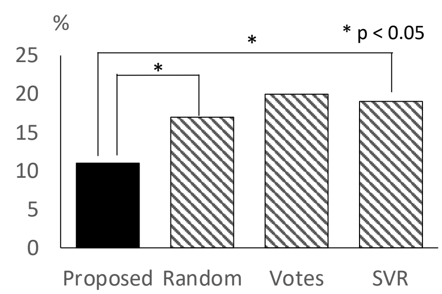

Fig. 7. Percentage of ”not sympathetic” reviews in top 10 ranked.

B. Experimental Results

We calculated the percentage of reviews that participants rated as “sympathetic” and the percentage of reviews that they rated as “not sympathetic” out of the top 10 ranked reviews in the proposed and comparative methods.

The results for the reviews evaluated as “sympathetic” are shown in Figure 6. When significant differences were confirmed by T-test, significant differences were observed between the proposed method and the random sampling method, and between the proposed method and the order of the number of votes, at a significance level of 5 percent.

The results for the reviews evaluated as “not sympathetic” are shown in Figure 7. When the T-test was used to confirm the significant differences, significant differences were observed between the proposed method and the random sampling method and between the proposed method and the support vector regression at a significance level of 5 percent.

C. Discussion

The top 10 reviews sorted based on the proposed method had the largest percentage of reviews rated as ”sympathetic” among all the methods shown in Figure 6. Significant differences were observed between the proposed method and the order of votes. Thus, it was found that the review ranking considering subjective preferences by the proposed method presented reviews that were more useful in terms of product selection at the top than the review ranking based on objective indices. The proposed method is more effective than the regression-based review recommendation since there is no significant difference between the regression method and the order of the number of votes. In addition, among all the methods shown in Figure 7, the proposed method had the lowest percentage of reviews that were input with a rating of “not sympathetic.” There was a significant difference between the proposed method and the support vector regression, indicating that the proposed method effectively filters reviews that are not useful for the target user in determining the product. The proposed method was found to be effective in filtering out reviews that are not useful for the target user’s evaluation. The percentage of reviews for which the rating of ”do not agree” was entered was significantly different as a percentage between the proposed method and the order of the number of votes but was not significantly different. The reason for this may be the small number of participants, which could be improved by improving the number of participants.

VI. Conclusion

This paper proposed a method to assist users in efficiently selecting items for recommendation by collaborative filtering on online review sites, focusing on the act of selection expected to occur after recommended items are presented to the user. To predict the value of each reviewer in the target user, the method uses the feedback of ratings on reviews of known items as input to predict the trust of the reviews as to whether the reviews are reviewer-like statements. The ranking of reviews is then based on the user's trust. To verify whether the review ranking sorted by the proposed method is helpful for users' product evaluation, we conducted a subject experiment using reviews on a reading meter. The experimental results showed that the proposed method is effective for users in ranking items because it gave higher priority to the reviews that the users could identify with and lower priority to the reviews that the users could not identify with.

As a future issue, we can conduct validation experiments using data from online review sites besides books. The proposed method can be adapted for books, movies, music, and other items because users' preferences for reviewers will likely differ. Therefore, we are considering developing a method that considers objective and subjective values. There is also a possibility that the proposed method can be adapted to social networking services and the sharing economy. Specifically, the proposed method could be applied to timeline filtering in Social Networking Services (SNSs), review recommendations in the sharing economy, and so on. Additionally, we are considering using advanced resource language models such as BERT [27] to more accurately predict the degree of empathy of unknown users.

References

[1] T. Ushiama, D. Minami, "Personalized Item Review Ranking Method Based on Empathy," In Proc. of The Sixteenth International Conference on Digital Society, 2022, pp. 42–43.

[2] C. D. Manning, P. Raghavan, and H. Schutze, *Introduction to Information Retrieval*, Cambridge University Press, 2008.

[3] M. L. Anderson and J. R. Magruder, "Learning from the crowd: Regression discontinuity estimates of the effects of an online review database," *Economic Journal*, vol. 122, issues 563, pp. 957–989, 2012.

[4] S. G. Esparza, M. P. O'Mahony, and B. Smyth, "Effective product recommendation using the real-time web," In Proc. of the 30th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, 2010, pp. 5–18.

[5] S. G. Esparza, M. P. O'Mahony, and B. Smyth, "A multi-criteria evaluation of a user-generated content based recommender system," In Proc. of the 3rd Workshop on Recommender Systems and the Social Web in RecSys'11, 2011, pp. 49–56.

[6] C.W.K. Leung, S.C.F. Chan, and F. Chung, "Integrating collaborative filtering and sentiment analysis: A rating inference approach," In Proc. of the ECAI 2006 Workshop on Recommender Systems, 2006, pp. 62–66.

[7] W. Zhang, G. Ding, L. Chen, C. Li, and C. Zhang, "Generating virtual ratings from chinese reviews to augment online recommendations," *ACM Trans. Intell. Syst. Technol*, vol. 4, no. 1, 2013.

[8] D. Poirier, F. Fessant, and I. Tellier, "Reducing the cold-start problem in content recommendation through opinion classification," In Proc. of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2010, pp. 204–207.

[9] C.C. Musat, Y. Liang, and B. Faltings, "Recommendation using textual opinions," In Proc. of the 23rd International Joint Conference on Artificial Intelligence (IJCAI'13), 2013, pp. 2684–2690.

[10] J. McAuley and J. Leskovec, "Hidden factors and hidden topics: Understanding rating dimensions with review text," In Proc. of the 7th ACM International Conference on Recommender Systems (RecSys'13), 2013, pp. 165–172.

[11] Y. Seroussi, F. Bohnert, and I. Zukerman, "Personalised rating prediction for new users using latent factor models," In Proc. of the 22nd ACM Conference on Hypertext and Hypermedia (HT'11), 2011, pp 47-56.

[12] Y. Wang, Y. Liu, X. Yu, "Collaborative filtering with aspect-based opinion mining: A tensor factorization approach," In Proc. of the IEEE International Conference on Data Mining (ICDM'12), 2012, pp. 1152–1157.

[13] H. Liu, J. He, T. Wang, W. Song, and X. Du, "Combining user preferences and user opinions for accurate recommendation," *Electron. Commer. Res. Appl.*, vol. 12, no. 1, 2013, pp.14–23.

[14] L. Chen and F. Wang, "Preference-based clustering reviews for augmenting e-commerce recommendation," *Knowl. Based Syst.*, vol. 50, pp. 44–59, 2013.

[15] A. Levi, O. Mokryn, C. Diot, and N. Taft, "Finding a needle in a haystack of reviews: Cold start context-based hotel recommender system," In Proc. of the 6th ACM International Conference on Recommender Systems (RecSys'12), 2012, pp. 115-122.

[16] L. Chen, G. Chen, and F. Wang, "Recommender systems based on user reviews: the state of the art," *User Modeling and User-Adapted Interaction*, vol. 25, pp 99-154, 2015.

[17] T. Hayashi and R. Onai, "Movie Recommendation Using Reviews on the Web," *Transactions of the Japanese Society for Artificial Intelligence*, vol. 30, no. 1, pp. 102–111, 2015.

[18] A Statistical Analysis of 1.2 Million Amazon Reviews: http://minimaxir.com/2014/06/reviewing-reviews, [Accessed June 1, 2023].

[19] S. Moghaddam, M. Jamali, and M. Ester. ETF, "Extended Tensor Factorization model for personalizing prediction of review helpfulness," In Proc. of the fifth ACM international conference on Web search and data mining (WSDM'12), 2012, pp. 163–172.

[20] S. Raghavan, S. Gunasekar, J. Ghosh, "Review quality aware collaborative filtering," In Proc. of the 6th ACM Conference on Recommender systems, 2012, pp. 123–130.

[21] A. Mukherjee, B. Liu, and N. Glance, "Spotting Fake Reviewer Groups in Consumer Reviews," In Proc. 21st International Conference on World Wide Web, 2012, pp. 191–200.

[22] S. Xie, G. Wang, S. Lin, and P.S. Yu, "Review SpamDetection via Temporal Pattern Discovery," Proc. 18th ACM International Conference on Knowledge Discovery and Data Mining, 2012, pp. 823–831.

[23] G. Wang, S. Xie, B. Liu, and S. Yu, "Review Graph based Online Store Review Spammer Detection," In Proc. 11th IEEE International Conference on Data Mining, 2011, pp. 1242–1247.

[24] Dokusho Meter, http://bookmeter.com/, [Accessed June 1, 2023].

[25] D. Minami and T. Ushiama, "Can you trust the user?: Collaborative trust estimation model for recommendations," In Proc. 2017 Twelfth International Conference on Digital Information Management, 2017, pp. 252–256.

[26] Q.L. Le and T. Mikolov, "Distributed Representations of Sentences and Documents," In Proc. of The 31st International Conference on Machine Learning, 2014, pp. 1188–1196.

[27] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," In Proc. of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, vol. 1, pp. 4171–4186, 2019.