Personalized Item Review Ranking Method Based on Empathy

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—In e-commerce, online reviews posted about items play an essential role in helping users select products. However, when many reviews are posted for the same product, it is sometimes difficult for users to find the most valuable reviews among them. This paper proposes a method for ranking online reviews of a target item based on the user's empathy for reviewers. Using the target user's feedback on reviews for known items as input, the proposed method estimates the empathy toward the reviewer and ranks reviews for unknown items based on it. Evaluation experiments showed that the proposed method is effective against comparative methods.

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

I. Introduction

E-commerce sites often provide reviews about products. Reviews play an important role in users' selection of items. However, when a large number of reviews about the same item are included, it is impractical to browse through all reviews [1]. 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.

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 user. Since existing review ranking 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 ranking online reviews of a target item based on user empathy. The proposed method predicts a target user's empathy for a reviewer based on the reviews the user has rated in the past. It then ranks the reviews based on their empathy for the reviewer.

Section II introduces a method for ranking item reviews based on empathy. Section III shows the experimental results for evaluating the effectiveness of the proposed method. Section IV describes the summary and future work.

II. Proposed Method

A. Typicality of reviews to authors

Reviews are considered to reflect the values and personality of the author (reviewer). The degree to which a review reflects the values and personality of the reviewer who wrote it is called the "typicality" of the review. $\text{typi}(d, r)$, which is the degree of typicality of a review document d to a reviewer r, is defined by the following equation.

$$\text{typi}(d, r) = p(r|d) \tag{1}$$

where $p(r|d)$ is the conditional probability that given a review document d, its author is d.

The proposed method uses machine learning to predict $p(r|d)$. Specifically, reviews are vectorized using Doc2Vec [2], and machine learning are performed with that author as the correct label. A neural network with two hidden layers is used for machine learning.

B. Similarity of reviewers

To predict empathy for unknown reviewers, similarities between reviewers are used. In a two-layer neural network that takes reviews as input and predicts authors, the all-connected layer in front of the output layer is considered to be the manifestation of the features used to classify reviewers. The vector obtained by inputting a single document that is a concatenation of all reviews posted by reviewer r as input to the machine learning model of author estimation is denoted as $\boldsymbol{v}(r)$.

$\text{sim}(r_1, r_2)$, which is the similarity between two reviewers r_1 and r_2, is defined by the following formula.

$$\text{sim}(r_1, r_2) = \frac{\boldsymbol{v}(r_1) \cdot \boldsymbol{v}(r_2)}{\|\boldsymbol{v}(r_1)\| \times \|\boldsymbol{v}(r_2)\|} \tag{2}$$

where $PR(u)$ represents the review set that was positively rated by user u, $NR(u)$ represents the review set that was negatively rated by user u and $\text{reviewer}(d)$ represents the reviewer who submitted a review d.

C. Estimation of the empathy for reviewers based on review ratings

The proposed method uses the user's ratings of previous reviews to extract the level of empathy toward the reviewer. In this study, we assume that users can vote "like" and "dislike" for some reviews. The user is not required to rate all reviews. We predict the empathy $\text{empathy}(u, r)$ of a user u toward a reviewer r who has posted a review he has rated, using the following formula.

$$\text{empathy}(u, r) = \sum_{d \in D_{pos}(r)} \text{typi}(d, r) - \sum_{d \in D_{neg}(r)} \text{typi}(d, r) \tag{3}$$

where $D_{pos}(r)$ represents the set of review documents posted by reviewer r that the target user rated as "like" and $D_{neg}(r)$ represents the set of review documents that the target user rated as "dislike". $\text{tipi}(d, r)$ denotes the topicality of a review document d to reviewer r.

D. Ranking of reviews for unknown items

Estimate the value of a review of an unknown item I_u based on the feedback of the evaluation of a review of a known item I_k. The recommendation score $score(u, d)$ of a review d for a target user u is defined by the following equation.

$$score(u, d) = \frac{\sum_{r \in R(u)} \text{empathy}(u, r) \times \text{sim}(r, \text{reviewer}(d))}{\sum_{r \in R(u)} \text{sim}(r, \text{reviewer}(d))}$$

(4)

The reviewer set $R(u)$ with respect to user u is defined by the following equation.

$$R(u) = \{x | x = \text{reviewer}(d), d \in PR(u) \cup NR(u)\} \quad (5)$$

where $\text{reviewer}(d)$ denotes the reviewer who submitted the review d.

III. EVALUATION

A. Experimental setup

Experiments were conducted to evaluate the effectiveness of the proposed method. The dataset used for the experiments was obtained by crawling from one of the famous Japanese online book review sites "Dokusho meter". Six subjects 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 subjects 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 of "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.

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.

B. experimental results

We calculated the percentage of reviews that subjects 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 that were evaluated as "sympathetic" are shown in Fig. 1. When significant differences were confirmed by T-test, significant differences were observed between the proposed method and the random sampling method,
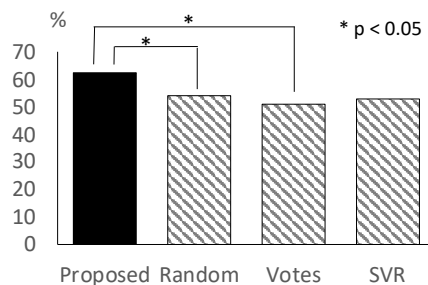

Fig. 1. Percentage of "sympathetic" reviews in top 10 ranked.
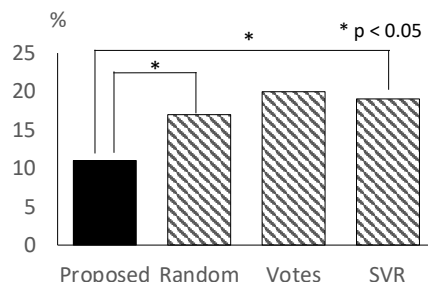

Fig. 2. Percentage of "not sympathetic" reviews in top 10 ranked.

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 that were evaluated as "not sympathetic" are shown in Fig. 2. 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.

The above results show that the proposed method includes more reviews with content that users can relate to and fewer reviews that users cannot relate to in the top ranking than the comparative methods.

IV. CONCLUSION

In this paper, we propose a review ranking method based on user empathy. The proposed method predicts the reviewer's empathy based on the user's review ratings to achieve a personalized ranking. Results of evaluation experiments using subjects showed the effectiveness of the proposed method. As future work, we are considering the use of advanced resource language models such as BERT [3] to more accurately predict the degree of empathy of unknown users.

REFERENCES

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

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

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