A Product Recommendation System Based on User Complaint Analysis Using Product Reviews

Terutaka Yoshikawa
Yamaguchi University, Japan
b091vg@yamaguchi-u.ac.jp

Yuanyuan Wang
Yamaguchi University, Japan
y.wang@yamaguchi-u.ac.jp

Yukiko Kawai
Kyoto Sangyo University, Japan
Osaka University, Japan
kawai@cc.kyoto-su.ac.jp

Abstract—In recent years, e-commerce attracts attention to the spread of online shopping sites, many users who refer to product review comments to choose the products and higher-rated products can be recommended from users' purchase histories. However, traditional recommendation systems cannot recommend alternative products when existing products' reviews contain user complaints (e.g., price, battery, design, function, etc), it is still difficult to satisfy users' requirements. Therefore, in this work, we propose a novel product recommendation system that analyzes two kinds of information: complaint information and satisfaction information from review comments on e-commerce. It is possible to recommend alternative products to satisfy your requirements that can solve the complaint information on the product when you browsing. In this paper, we describe the complaint information extraction based on product review analysis by extracting both negative information and positive information from product reviews, and we also explain alternative product recommendation method that can solve complaints, and we verify the effectiveness of complaint information extraction and alternative product recommendation.

Index Terms—e-commerce, review analysis, complaint information, product recommendation

I. INTRODUCTION

In recent years, e-commerce online shopping sites are widely used for many users such as Amazon[1] and Rakuten Ichiba[2], and they help users to find suitable products easily. Review comments are an increasingly important part of the purchase journey for online consumers, they provide helpful information for user decision-making. Many e-commerce recommendation systems have been proposed based on review comments by researchers [1]–[4]. These systems help users to find products and promote their purchase. In addition, users can refer to other consumers' reviews to acquire information about a target product before purchasing. However, many products have both positive reviews with high ratings and negative reviews with low ratings. In these cases, users often cannot recognize genuine complaints about the target product, since negative reviews with low ratings tend to be ignored. Moreover, even users find reliable information, they still difficult to find products that meet their requirements.

Therefore, in this work, we propose a novel product recommendation system to facilitate users to find suitable alternative products that will address complaints regarding products the

[1] https://www.amazon.com/
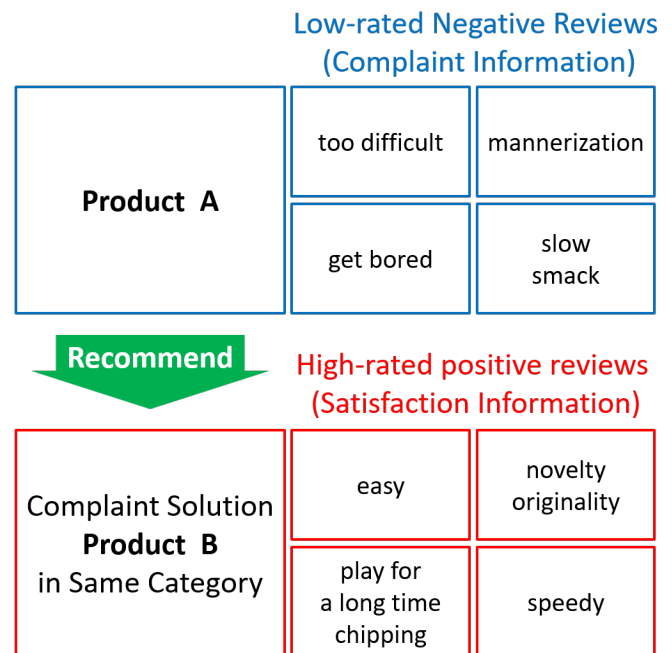[2] https://www.rakuten.co.jp/

Fig. 1. A complaint solution product recommendation system by analyzing product reviews.

users are browsing by extracting both negative and positive reviews (see Fig. 1). This system has 3 components: 1) product review analysis by acquiring reviews of products and classifying the reviews into low-rated reviews and high-rated reviews; 2) negative (complaint) information and positive (satisfaction) information extraction by determining negative feature words with low ratings and positive feature words with high ratings; 3) alternative product recommendation by calculating the similarities between negative (complaint) information and positive (satisfaction) information. It is possible to recommend alternative products with positive reviews with high ratings which solve complaints of products with negative reviews with low ratings.

The remainder of this paper is structured as follows. The next section provides a summary of related work. Section III describes the analysis method of product reviews and explains how to recommend complaint solution products. Section IV

shows our evaluation experiments of the proposed method using the real dataset. Finally, Section V concludes this paper and presents future works.

II. RELATED WORK

Since e-commerce sites became popular, a lot of product recommendation systems are proposed. Yandi et al. [7] proposed a product recommendation system by applying coupons. They supposed that consumers' purchasing is related to the behavior when they choose coupons. Vivek et al. [8] proposed a system which recommends products with suitable size. The recommended products are learned from user history data using latent factor model which are widely used for product review analysis [9], [10]. In this work, we propose a complaint solution product recommendation systems based on user complaint analysis. In addition, different methods of recommending alternative products have been studied and are well-documented. McAuley et al. [11] considered the relationships between alternative products based on products with reviews and their costs. Their proposed method for clarifying the relationship between alternative products for a given topic. Zheng et al. [12] also analyzed the relationships between alternative products by applying economic principles. It is similar to our proposed system that recommends alternative products that can solve users' complaints on their target products.

Several studies on review analysis are popular as well. Dong et al. [5] proposed an attention-enhanced attribute-to-sequence model to generate product reviews for given attribute information. McAuley et al. [13] analyzed user experiences from reviews. Koren [14] analyzed the movement of users interests from reviews. Yong et al. [15] measured bias users from reviews. They generated preference vectors and evaluation vectors to analyze each review. In this work, we use product reviews to analyze features of items and to extract complaint information and satisfaction information for item recommendation.

Review bias has also been extensively studied. Zhang et al. [16] calculated review bias using two factors: user preferences and prejudices caused by reading other reviews. Since we propose a product recommendation system based on review data, we will analyze the reliability of said data. Internet complaints are also discussed in some research. Before SNS came out, Jean [17] already implied complaints on Internet forum could be used for user analysis for the company. Mitsuzawa et al. [6] analyzed complaint data provided by Insight Technology, Inc. from the Fuman Kaitori Center website, which provides a platform for consumers' complaints. They described the feature of complaint data posted on the platform they made. Hasegawa et al. [18] used the same complaint data to visualize complaint objects classified by occupations and ages. Suehiro et al. [19] characterized the same complaint data and analyzed the semantic information by applying the Bag-of-Words model. Hayashi et al. [20] proposed a product recommendation method for solving users' complaints using both complaint data and review data. Our goal is similar to it, we propose a product recommendation method for solving users' complaints by extracting both complaint information and satisfaction information from product reviews.

III. COMPLAINT SOLUTION PRODUCT RECOMMENDATION BASED ON REVIEW ANALYSIS

A. Extracting Complaint Information

In this study, we analyze product reviews and extract complaint information for each product. Products with a rating of 3 or less are regarded as low evaluation ($\leq$ 3 stars), and their reviews are regarded as low-rated reviews. However, some low-rated reviews contain high evaluation content. For example, a negative review "I was looking forward to it before I bought it, but I was disappointed when I actually tried it.", if the phrase "looking forward" has a positive meaning, it would be excluded from the negative reviews. The positive information should be excluded from negative reviews. In addition, the negative information in the low-rated review is regarded as complaint information. In this paper, we determine the positive and negative meanings for each word by using the Japanese Sentiment Polarity Dictionary[3], which is an open resource published by Tohoku NLP laboratory[4]. We calculate the weight of each feature word wn expressing complaints of a product using the following *TF-IDF* formula.

$$TF = \frac{\#wn \text{ in complaint information of a product}}{\text{total \#words in reviews of a product}} \quad (1)$$

$$IDF = \log \frac{\text{total \#reviews of a product}}{\#\text{reviews of a product contain } wn} \quad (2)$$

Then, the feature vector of complaint information is generated as $v_F = (wn_1, \cdots, wn_i, \cdots, wn_m)$ by the calculated weight of each feature word. Here, *TF* returns the word frequency of complaint information of a target product. *IDF* returns inverse document frequency when each review of the target product is considered as one document.

B. Extracting Products with Positive Evaluation

Products with a rating of 4 and 5 are regarded as high evaluation ($>$ 3 stars), and their reviews are regarded as high-rated reviews. However, high-rated reviews may include low evaluation content. For example, a review "XX was not good, but ○○ was good.". The negative information should be also excluded from positive reviews. In addition, the positive information in the high-rated review is regarded as satisfaction information. Then, the complaint solution information corresponding to the complaint information is extracted from the feature words. In order to determine the positive and negative meanings for each word, we use the Japanese Sentiment Polarity Dictionary in Section III-A. We calculate the weight of each feature word wp expressing satisfaction information of a product using Eqs.1 and 2 as in Section III-A. Then,

---

[3]http://www.cl.ecei.tohoku.ac.jp/index.php?Open%20Resources%2FJapanese%20Sentiment%20Polarity%20Dictionary
[4]https://www.nlp.ecei.tohoku.ac.jp/

TABLE I
FIVE PRODUCTS USED IN OUR EVALUATION

No.	Product Title
A	New Super Mario Bros. Wii (Nintendo)
B	Wii Sports Resort (Nintendo)
C	Taiko no Tatsujin Wii Party with Everyone 3rd generation! (BANDAI NAMCO Games)
D	Momotaro Dentetsu 2010: Sengoku Ishin no Hero Daishugo! No Maki (Hudson)
E	Kirby's Return to Dream Land (Nintendo, HAL Laboratory)

TABLE II
EXTRACTED COMPLAINT INFORMATION OF THE PRODUCTS

Product No.	Top 20 Feature Words of Extracted Complaint Information
A	look, busy, overwork, expensive, what, difficult, cancel, child, first, bored, selfishness, forgetful, unforgettable, unskillful, wasteful, enduring, fighting, bumping, senseless, distracting
B	worst, price, bored, very, bad, disgusting, failure, used, improper, impulsive, tired, wait, poor, child, problem, stress, embarrassment, extreme, lack, regret
C	regret, minimum, nothing, break, disappoint, bad, bored, child, degree, later, sold out, panicked, amount, wasteful, worst, complaint, repulsion, cancellation, used
D	missteps, means, price, lukewarm, slow, child, evil, collapsing, coercion, Binbogami, reflect, degree, Normal, problem, conservative, poor, no, detour, perseverance
E	ignore, price, delay, complaint, annoyance, no, failure, not enough, look, special, limited, enemy, weird, easy, inconvenient, complain, busy, possible, child, remove

the feature vector of satisfaction information is generated as $v_F = (wn_1, \cdots, wn_i, \cdots, wn_m)$ by the calculated weight of each feature word.

C. Recommending Complaint Solution Products

In order to recommend alternative products to solve complaints of products when you are browsing, it is necessary to compare products in the same categories between their complaint (negative) information and satisfaction (positive) information. This section describes the similarity calculation of complaint feature vectors of target products and satisfaction feature vectors of other products in the same categories obtained in the previous section. We calculate the similarities of complaint feature vectors of target products and satisfaction feature vectors of other products by using the cosine similarity as follows:

$$Sim(v_F, v_R) = \frac{\sum wn_m \cdot wp_n}{\sqrt{\sum (wn_m)^2} \cdot \sqrt{\sum (wp_n)^2}}$$

Here, v_F denotes a negative feature vector of complaint information for each low-rated product calculated from its reviews, and v_R denotes a positive feature vector of satisfaction information for each high-rated product calculated from its reviews. In this work, in order to recommend products to solve the complaints, v_F and v_R are required to be in the same categories. Firstly, in order to calculate the cosine similarity, the dimension numbers of two kinds of feature vectors are aligned in the same categories. Secondly, we normalize the *TF-IDF* values of these feature vectors. Therefore, the satisfaction feature vectors of the product which has high similarity value with the complaint feature vectors can solve their complaints.

IV. EVALUATION

In our evaluation, we used product data and review data from "Rakuten Dataset" provided by Rakuten, Inc. provided by IDR Dataset Service of National Institute of Informatics (NII). It has approximately 156 million products and 64 million reviews from August 4, 2010 to April 1, 2014. In this work, we extract negative (complaint) information and positive (satisfaction) information from reviews of each product. Because of several Japanese words are the same meaning written in Japanese Kanji characters or hiragana, we selected the words with a higher *TF-IDF* value. We evaluated five products (A,

TABLE III
TOP 5 COMPLAINT SOLUTION PRODUCTS

Product No.	Product Titles of Top 5 Recommended Products (Cosine Similarity)
A	SD Gundam Generation Wars (0.693) Dragon Ball: The Great Adventure (0.681) Super Smash Bros. Brawl (0.654) Super Mario Galaxy 2 (0.653) Super Mario Collection Special Pack (0.652)
B	Famista DS 2010 (0.716) Inazuma Eleven 3: Challenge to the world!! spark (0.706) Mario & Sonic at the Vancouver Olympics (0.698) Wii Fit Plus (0.694) Family Fishing (0.675)
C	Taiko no Tatsujin Tiny Dragon and Wonderful Orb (0.701) Taiko no Tatsujin Wii definitive edition (0.699) JUST DANCE Wii (0.698) Taiko no Tatsujin Wii Super strong version (0.671) Everyone's Rhythm Heaven (0.669)
D	Puyo Puyo 7 Special Price Wii Version (0.748) THE Table Game Mahjong, Igo, Shogi, Card, Reversi, Gomoku Narabe (0.723) Momotaro Dentetsu 16 Hokkaido Daikotomaki! Everyone's recommended selection (0.712) Karaoke JOYSOUND Wii SUPER DX (0.694) Itadaki Street Wii (0.693)
E	Mario Cart Wii (0.766) Super Smash Bros. Brawl (0.759) Kirby's Epic Yarn (0.757) Super Mario Galaxy 2 (0.720) Monster Hunter 3 (0.712)

B, C, D, E), each product had a particularly high number of review comments (see Table I).

A. Experiment I: Verification of Complaint Information Extraction

In Experiment I, we verify the extracted complaint information on the products in Table I. The complaint information is shown as the top 20 negative feature words based on their *TF-IDF* values in Table II, and the complaint feature words are underlined. From the result of extracted complaint information, we found that there were many reviews about family games mainly on parents and children. Also, we confirmed that we could extract feature words of complaint information on products.

B. Experiment II: Verification of Complaint Solution Product Recommendation

In Experiment II, we verify the recommended products which solve the complaints of each product in Table I. The top 5 recommended products in the same categories with a high cosine similarity of feature words in each product are shown in Table III. As a result, the same series was ranked high, and the new series was able to solve the complaints of earlier products. Other sports products ranked high for physical games (more commonly called sports, to be more specific), while party games ranked high for multiplayer games (two or more people playing a game on the computer and competing with each other over the Internet or in-house local network).

V. CONCLUSION

In this paper, we proposed a novel product recommendation system to solve users' complaints on products by analyzing both complaint information and satisfaction information from product reviews. We verified the effectiveness of our proposed complaint information extraction method and our proposed complaint solution product recommendation method.

In the future, we will improve our complaint information extraction method. Since the complaint information in the e-commerce often included the procedure of the dispatch, we should remove the complaint information for shopping sites or delivery companies. In addition, we also plan to evaluate our product recommendation system a questionnaire survey.

ACKNOWLEDGMENT

REFERENCES

[1] Li Chen, Guanliang Chen, Feng Wang: Recommender Systems Based on User Reviews: The State of the Art, User Modeling and User-Adapted Interaction, vol. 25, no. 2, pp. 99–154, 2015.

[2] J. Ben Schafer, Joseph A. Konstan, John Riedl: E-Commerce Recommendation Applications, Data Mining and Knowledge Discovery, vol. 5, no. 1-2, pp. 115–153, 2001.

[3] Brent Smith, Greg Linden: Two Decades of Recommender Systems at Amazon.Com, IEEE Internet Computing, vol. 21, no. 3, pp. 12–18, 2017.

[4] Qi Zhao: E-commerce Product Recommendation by Personalized Promotion and Total Surplus Maximization, Proceedings of the 9th ACM International Conference on Web Search and Data Mining (WSDM 2016), pp. 709–709, 2016.

[5] Li Dong, Shaohan Huang, Furu Wei, Mirella Lapata, Ming Zhou, Ke Xu: Learning to Generate Product Reviews from Attributes, Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, vol. 1, pp. 623-632, 2017.

[6] Kensuke Mitsuzawa, Maito Tauchi, Mathieu Domoulin, Masanori Nakashima, Tomoya Mizumoto: FKC Corpus: a Japanese Corpus from New Opinion Survey Service, Proceedings of the Novel Incentives for Collecting Data and Annotation from People: types, implementation, tasking requirements, workflow and results, pp. 11-18, 2016.

[7] Yandi Xia, Giuseppe Di Fabbrizio, Shikhar Vaibhav, Ankur Datta: A Content-based Recommender System for E-commerce Offers and Coupons, Proceedings of the SIGIR 2017 Workshop on eCommerce (ECOM 2017), 7 pages, 2017.

[8] Vivek Sembium, Rajeev Rastogi, Atul Saroop, Srujana Merugu: Recommending Product Sizes to Customers, Proceedings of the 11th ACM Conference on Recommender Systems (RecSys 2017), pp. 243–250, 2017.

[9] Zhiyong Cheng, Ying Ding, Lei Zhu, Mohan Kankanhalli: Aspect-Aware Latent Factor Model: Rating Prediction with Ratings and Reviews, Proceedings of the 2018 World Wide Web Conference (WWW 2018), pp. 639–648, 2018.

[10] Zhu Sun, Guibing Guo, Jie Zhang, Chi Xu: A Unified Latent Factor Model for Effective Category-Aware Recommendation, Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization (UMAP 2017), pp.389–390, 2017.

[11] Julian McAuley, Rahul Pandey, Jure Leskovec: Inferring Networks of Substitutable and Complementary Products, Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2015), pp. 785–794, 2015.

[12] Jiaqian Zheng, Xiaoyuan Wu, Junyu Niu, Alvaro Bolivar: Substitutes or Complements: Another Step Forward in Recommendations, Proceedings of the 10th ACM conference on Electronic commerce (EC 2009), pp. 139–146, 2009.

[13] Julian John McAuley, Jure Leskovec: From Amateurs to Connoisseurs: Modeling the Evolution of User Expertise Through Online Reviews, Proceedings of the 22Nd International Conference on World Wide Web (WWW 2013), pp. 897–908, 2013.

[14] Yehuda Koren: Collaborative Filtering with Temporal Dynamics, Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2009), pp. 447–456, 2009.

[15] Yong Ge, Jingjing Li: Measure and Mitigate the Dimensional Bias in Online Reviews and Ratings, Proceedings of the 2015 International Conference on Information Systems (ICIS 2015), pp. 445–455, 2015.

[16] Xiaoying Zhang, Junzhou Zhao, John C.S. Lui: Modeling the Assimilation-Contrast Effects in Online Product Rating Systems: Debiasing and Recommendations, Proceedings of the 11th ACM Conference on Recommender Systems (RecSys 2017), pp. 98–106, 2017.

[17] L. Jean Harrison-Walker: E-complaining: A Content Analysis of an Internet Complaint Forum, Journal of Services Marketing, vol.15, no.5, pp.397–412, 2001.

[18] Toru Hasegawa, Daisuke Kitayama: Visualization of Complaint Groups using Complaint Data, Proceedings of the 9th Forum on Data Engineering and Information Management (DEIM Forum 2017), P7–1, 2017.

[19] Shun Suehiro, Hiroaki Saito: Feature Vectorization of Complaint Data Sets, Proceedings of the 23th annual meeting of the Association for Natural Language Processing (NLP 2017), pp. 545–548, 2017.

[20] Toshinori Hayashi, Yuanyuan Wang, Yukiko Kawai, Kazutoshi Sumiya: An E-Commerce Recommender System using Complaint Data and Review Data, Proceedings of ACM IUI2018 Workshop on Web Intelligence and Interaction (WII 2018), vol. 2068, no. wii8, pp. 1–5, 2018.