Proposal of a Cosmetic Product Recommendation Method with Review Text that is Predicted to Be Write by Users

Natsumi Baba, Yuichi Sei[a], Yasuyuki Tahara[b] and Akihiko Ohsuga[c]
The University of Electro-Communications, Tokyo, Japan

Keywords: Collaborative Filtering, Morphological Analysis, Review Analysis, Product Recommendation.

Abstract: There are a variety of product introduction sites on the Internet, and many of these usually provide a combination of product composition information and user review text. It is difficult to understand the features of a product in detail from the information on these sites. Furthermore, these review sites often include product recommendations such as "recommended for you," but often lack an explanation of why the product is recommended. Therefore, this study proposes an approach that provides both the user's opinion of the product and the reason for recommending the product in a simplified manner. Using cosmetics as a case study, where the user's actual experience is important, we scored product features on a 5-point scale based on review submitted by users. This data was used for collaborative filtering to determine product recommendations and generate review sentences that target users are expected to write when using the product. The generated reviews facilitate users to understand the details of a product before purchasing it and are useful for comparison before purchasing a product. To verify the usefulness of the proposed method, we conducted a questionnaire comparing it with existing methods. The proposed method aims to improve user satisfaction in product recommendations.

1 INTRODUCTION

In recent years, there have been many product reviews on Social Networking Service and video sharing sites, and many people refer to Internet review information when purchasing products. Cosmetics are goods for which reviews are particularly important, since there is a large difference in suitability for different people, and information on the type of person for whom the product is suitable is important. On many review sites, product reviews are composed of a score that represents the overall evaluation of the product and a free description by the reviewer. Although there are many sites where product reviews are posted, most of them consist only of information on the product components, the average value of the overall rating obtained for the product, and the actual product review. Therefore, it is difficult to judge whether a product is worth buying or not before purchasing it.

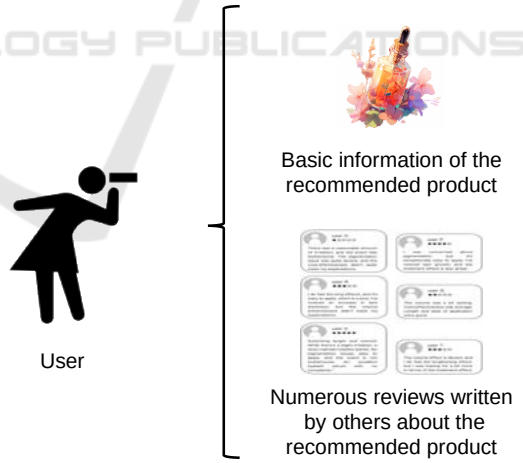

Figure 1: General process of product selection. It is necessary to carefully check numerous reviews of many similar products.

Product recommendation methods using reviews have long been studied for a variety of goods

[a] https://orcid.org/0000-0001-6717-7028
[b] https://orcid.org/0000-0002-1939-4455
[c] https://orcid.org/0000-0002-2552-6717

Baba, N., Sei, Y., Tahara, Y. and Ohsuga, A.
Proposal of a Cosmetic Product Recommendation Method with Review Text that is Predicted to Be Write by Users.
DOI: 10.5220/0012377500003636

In *Proceedings of the 16th International Conference on Agents and Artificial Intelligence (ICAART 2024) - Volume 3*, pages 609-616
ISBN: 978-989-758-680-4; ISSN: 2184-433X

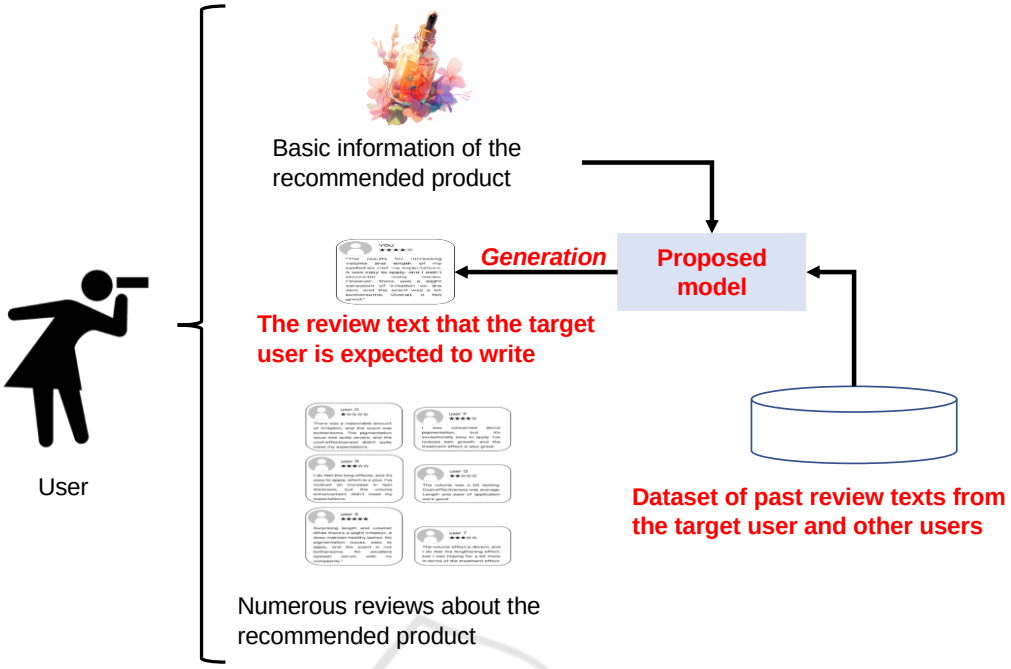

Figure 2: Proposed process of product selection. The user can understand in advance how she will feel by using the recommended product.

(Kirubanantham et al., 2022) (Zarzour et al., 2022) (Chehal et al., 2022) (Ambolkar at al., 2022). Among the goods, a number of methods have been proposed, including product recommendation methods specific to cosmetic product reviews (Iwabuchi et al., 2017) (Nakajima et al., 2019) (Sahar et al., 2022).

To solve the problem of not being able to easily grasp the features of a product and whether it is suitable for oneself as shown in figure 1, this study scoring products for each evaluation feature based on the review text and product score assigned to the reviews. Then, generates review text indicating how the user feels about the recommended product. Finally, achieve the situation that people can grasp the features of a product as shown in Figure 2.

This paper is organized as follows. Section 2 describes the research background of product recommendation using review text analysis and other methods. Section 3 describes the proposed method, and Section 4 describes the experiments and results. Then, Section 5 describes the possible implications of each result. Finally, conclude this paper and discuss future issues.

2 RELATED WORK

Matsunami et al. (2016) reported a study that conducted scoring by evaluation features using text reviews. They extracted evaluation phrases from reviews about "lotion," scored them manually for each evaluation feature, divided them, and constructed a dictionary of evaluation phrases based on co-occurrence of keywords. They proposed a method for automatically scoring each feature of a given text review by counting the evaluation features and their scores if any of the sentences containing the keywords satisfy the co-occurrence conditions described in the evaluation expression dictionary. Since this automatic scoring also takes into account the evaluation of specific evaluation items that are not directly related to the product, they conclude that countermeasures against noise expressions are necessary.

As a study using a dictionary of evaluation expressions, they describe the work of Taniguchi et al. (2019). They proposed the following method for efficiently creating evaluation expression dictionaries, which had been manually created by Matsunami et al. The first method is to use an existing evaluation expression dictionary. However, this method can be used only when the similarity of products such as "lotion" and "milky lotion" is high. It is found to be of limited use. Therefore, they proposed a method for automatically extracting candidate keywords when constructing the evaluation expression dictionary. In this method, TF-IDF values of nouns obtained by morphological analysis of

review sentences are obtained, and the top 1% of the TF-IDF values are selected as candidate keywords. About 30% of the automatically extracted words were actually adaptable to keywords. Based on these candidate keywords, they created a dictionary of evaluative expressions using the same method as Matsunami et al.'s. They concluded that this method obtained a higher evaluation rate than existing methods, even though the number of registered evaluative expressions was much smaller than existing methods. Another study that automatically constructs a lexicon of evaluation expressions is that of Sakai et al. (Sakai et al., 2019). They used a TF-IDF method to determine keywords, then used a dependency analyzer to collect adjectives, adverbs, verbs, nouns, and auxiliary verbs that modify the keywords, and created a dictionary of evaluation expressions. Sakai et al.'s proposed method has shown that the scoring results for items such as "Lame is amazing", which can be perceived in various ways by different people, are significantly different from the results obtained by manual scoring.

Okuda et al. (2020) studied product recommendation using this evaluation expression dictionary. When "User A" gives a high rating to "Product 1", they score the reviews that "User A" gave to "Product 1", and recommend products that have similar ratings to "Product 1". They concluded that determining similar users and recommending products based solely on scoring results is not an optimal method, and that future work should take into account the bipolarity of scores by evaluation features, user attribute information, and other factors to calculate user similarity and improve accuracy.

Yabe et al. (2021) proposed a method in which the user sets a reference product, and the system adjusts the score of each item based on the scoring results of that product to recommend products that are closer to the user's ideal. The system proposed by Yabe et al. showed better results than conventional systems in various items such as the ease of obtaining product information and the ease of reflecting user preferences.

As an example of a product recommendation method, Hara et al. (2021) proposed an experiment in which the product itself is turned into an agent and product recommendations are made. Through the agentization of the product, some people felt as if they were being recommended by a store clerk, and changed their selection.

In addition, a method has been proposed to construct a product recommendation system using deep learning, and to provide the basis for the output of the recommendation system at the time of recommendation. For example, Onogawa et al. (2020) proposed a method that uses LIME to identify words that characterize each product when comparing products. Imafuku et al. (2021) proposed a similar method to improve recommendation effectiveness and satisfaction. Afchar et al. (2020) proposed a method that visualizes which features, which are input to deep learning, contribute to the recommendation results. These methods show the basis of recommendation by highlighting words or features, and do not infer how the user actually perceives the product to be recommended. In addition, LIME assumes that all observations are independent, so it cannot account for nonindependent observations with high fidelity (Matsushima et al., 2021). This study targets users who write many review sentences after purchasing and using a product, and although a product has various features, it is assumed that users will mainly write about features that are of particular interest to them. Therefore, the review text that is predicted to be written by the user is considered to be highly useful as a reason for product recommendation.

Based on the results of these studies, scoring user reviews, predicting recommended products from them, and generating the reviews that the users are predicted to write when purchasing this product would allow users to easily and accurately review products before purchasing them, thus enabling them to make low-risk purchases.

3 METHOD

The proposed method mainly consists of the following 4 steps. It is structured as shown in Figure 3. First, the evaluation expression dictionary is created based on the review ratings given to the products. Next, scoring is performed for each evaluation product. Then, based on the scoring results and the user's past review data, the system decides which products to recommend. Finally, the system presents review sentences that predict how the user will evaluate the product when he or she purchases it, along with the recommended product.

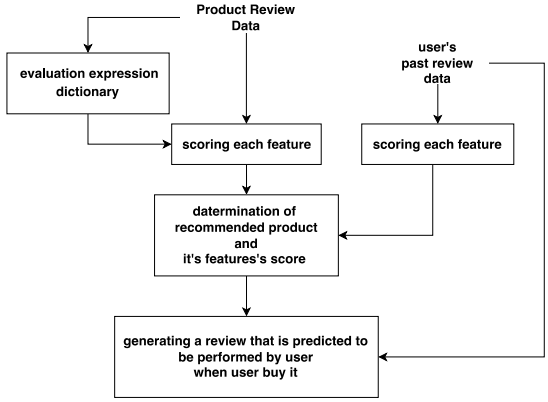

Figure 3: Schematic diagram of the proposed method.

As a method for scoring review texts for each evaluation product, an evaluation expression dictionary based on co-occurrence expressions, as in previous studies was used. The evaluation expression dictionary is created by first extracting evaluation expression phrases with reference to, performing scoring, and then dividing them to create an evaluation expression dictionary consisting of keywords, feature words, and words expressing degree.

Scoring of review texts is performed by morphological analysis of the review texts and checking whether feature words and words expressing degree appear in sentences containing words corresponding to the keywords in the evaluation expression dictionary. If a negative word is included in a sentence containing the keyword, the score of the evaluation expression feature is changed to the second value from the bottom if it was the second value from the top. All evaluation feature that were not mentioned in the review text were considered to be of standard satisfaction that did not need to be mentioned, and were given the median score. Based on the feature scores obtained, collaborative filtering was used to determine which products to recommend to the user.

The system estimates how users feel about the recommended product when they actually use it, based on other users' reviews, and generates the review sentences that users are expected to give based on the predicted scores. When generating review sentences, information on feature that users consider important when making a purchase, such as the number of times a product has been mentioned in past reviews, should be made easy to understand. Specifically, for a product that has a high overall satisfaction level but lacks the ease-of-use that the user values at 2 stars, this method generate a review sentence such as "Overall, I was very satisfied with the product, but I was disappointed that it was a little difficult to handle". The review sentence is generated as follows.

First, scoring of past product review sentences of the target user for each feature. Also, feature that the user pays particular attention to in the product are identified. Then, based on the score results, users who are assumed to be similar to the target user are calculated by collaborative filtering. Next, based on the ratings of similar users, calculate the number of scores for each feature in the product to be recommended and in the reviews that users are expected to give to that product. Finally, generate a review text that the user is expected to give when he/she purchases the recommended product by inputting the predicted score of each feature, information on the feature items that the user is interested in, and several past user reviews as prompts. See Chapter 4 for a concrete image of the proposed methodology.

4 EXPERIMENTS

We conducted a user study evaluation comparing a baseline method, which directly presents users with numerous reviews from other users, with our proposed method. It should be noted that the output of the proposed method consists solely of a single review that appears as if written by the user being recommended to. Thus, it is feasible to employ both the baseline and proposed methods concurrently. Consequently, even if fewer participants favored the proposed method over the baseline, it doesn't necessarily indicate the ineffectiveness of the proposed method. If even a modest number of participants find the proposed method preferable, it underscores the merit of appending the output of the proposed method alongside the baseline's output.

4.1 Experimental Setting

To create a dictionary of evaluation expressions for cosmetic items, this study decided to treat "eyelash serum", and conducted the following experiment using about 10,000 reviews of "Eyelash Serum" from Rakuten Group, Inc.'s review data (Rakuten Group, Inc., 2014).

First, obtained 316 evaluation phrases from 250 randomly selected reviews. Based on these, we obtained 678 co-occurrence expressions for evaluation expression dictionary. In this experiment, 8 features were set as evaluation deatures for eyelash serum: "Cost-Effectiveness", "Growth", "Volume",

"Hypoallergenic", "Treatment", "Pigmentation", "Ease of Application" and "Smell". The score given to each feature in this evaluation expression dictionary was also set to 1~5, in accordance with the score given to Rakuten's reviews.

Scoring of product evaluation features was performed on morphologically analyzed text reviews using the created evaluation expression dictionary.

Performing user-based collaborative filtering based on the scored ratings of each item and user information, and given data on product reviews that a user had performed in the past, it presented products that were highly rated by similar users and that the user was also likely to like.

In order to compare the proposed method with existing methods, such as the display method on actual product review sites, a questionnaire was conducted. with 19 men and women in their teens to 30s who often refer to reviews on online shopping sites. The questionnaire was conducted online, and respondents rated on a 5-point scale what they thought was the best way to present the questions and images that followed. The question consists of 3 parts. First part is Question 1, which compared the method of listing reviews of other companies commonly found on existing sites with the method of displaying reviews that are expected to be done by the user generated by the proposed method. Then, second part is Question 2, which compared the occasion of displaying only the score for each feature calculated by the proposed method and the occasion of generating and displaying reviews that are expected to be done by the user generated by the proposed method. Next, final part is Question 3, which tested whether weighting the review sentences to mention more features that the user is expected to place importance on would convey more details about the product.

4.2 Comparison of Others' Reviews and Your Own Style of Review

First, to compared the meåthod of listing reviews of other companies commonly found on existing sites with the method of displaying reviews that are expected to be done by the user generated by the proposed method, compared the method shown in Figure 4, which simply displays a list of other people's reviews, with the proposed method shown in Figure 5, which presents only the text of reviews generated by the proposed method. In this question, we refer to the former as “A” and the latter as “B”.

The results are shown in Figure 6. Those who answered that “A was easier to understand” were asked about additional reasons for their choice, and responses such as "B is easier to understand if you compare per review" and "I prefer to have reviews from various people”.

Figure 4: Displayed reviews in existing method.

Figure 5: Displayed reviews in proposed method.

4.3 Comparison of Features’ Score and Review Text Displays

To compared the occasion of displaying only the score for each feature calculated by the proposed method and the occasion of generating and displaying reviews that are expected to be done by the user generated by the proposed method, asked three questions to determine which method is better for understanding the feature of a product when purchasing it. There are two ways: one is to display only the score of each item obtained in the process of the proposed method, as shown in Figure 7, and the

other is to display the review text, as shown in Figure 5. In this question, the former is referred to as "A" and the latter as "B".

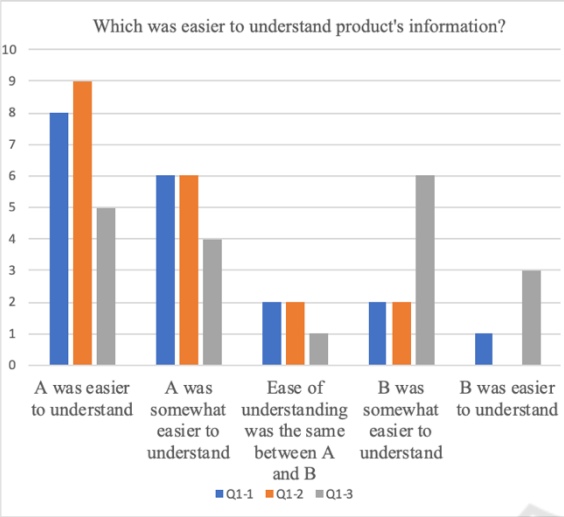

Figure 6: Result of Question1: “Which was easier to understand product’s information?”.

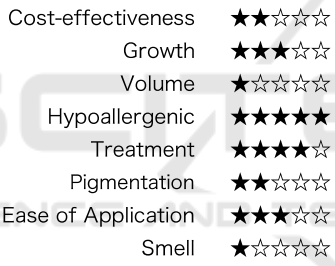

Figure 7: Displayed score of each feature in proposed method.

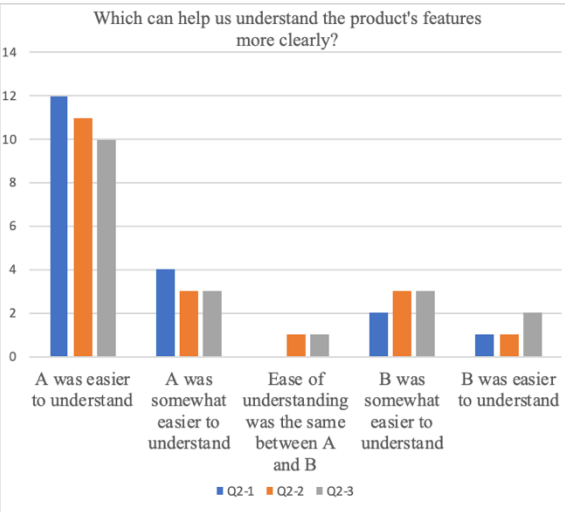

Figure 8: Result of Question 2: “Which can help us understand the product's features more clearly?”.

The results are shown in Figure 8. In all questions, the format displaying the numerical score was superior, but it was found that there were a certain number of people for whom the review text was easier to understand.

4.4 Weighting of Features

In the review text shown in Figure 5, 11 review texts, one in which only the score information of each item was used in generating the review text, we displayed and another in which the review text was weighted to mention more items that were considered important by the user. The question as to whether the information on each item was sufficiently conveyed from each review text was asked with respect to three items.

Table 1 shows the percentage of each item's characteristic information conveyed in the unweighted and weighted review sentences, respectively. In the case of the unweighted review sentences, the most salient features of the product were often mentioned and some information was not conveyed, but the weighted review sentences showed a significant improvement.

Table 1: Result of Question 3: Percentage of information conveyed for each feature (%).

	Hypoallergenic	Growth	Cost-Effectiveness
Weighted	52.63	78.95	94.74
Unweighted	0.00	10.53	0.00

5 DISCUSSION

The proposed method aims to improve user satisfaction with product recommendations by generating product recommendations using automatic scoring of review sentences. Based on this, discuss the results of the questionnaire survey conducted this time.

5.1 Comparison of Others' Reviews and Your Own Style of Review

This part, which compares the existing review sites with the proposed method, showed that the existing method is easier to understand. In this question, the number of reviews by others used to express the existing method was 6 for one product, which is easily readable, while in actual review sites, there are hundreds of reviews for a single product, and this is thought to be one of the reasons for the superiority of

A. This is thought to be one of the reasons for A's superiority. As shown in the additional questionnaire, personal preference is also considered to be an issue. In addition, in questions 1-3, the number of respondents who answered that B was easier to understand/slightly easier to understand than 1-1 and 1-2 was considerably larger, suggesting that the content of the displayed reviews may also be a factor. We would like to continue to examine the comparison with the existing method by changing the conditions, etc.

5.2 Comparison of Score and Text Displays

This part, which compared the score display only with the display in the review text, showed that the score display was easier to understand. However, in all questions, some respondents answered that the review text by several people B was easier to convey. Therefore, believe that the information may be more easily conveyed to users if the obtained scores and the predicted review sentences are listed together.

5.3 Weighting of Features

This part, which compares weighted and unweighted reviews, shows that weighting makes it easier to convey information about specific items. This weighting is thought to make it easier to convey information that is not often mentioned in ordinary reviews, but that the user cares about.

6 CONCLUSION

Based on the experimental results, the efficacy of generating and presenting a single review that appears as if written by the user being recommended to was confirmed. However, when compared to a vast number of reviews written by other users, its utility was found to be relatively lower. We believe that it can contribute to users' easy understanding of information when deciding whether to purchase a product by combining the scoring given when a product is purchased with review text tailored to the user's preferences.

As future prospects for this research, we would like to improve the usefulness of this method in comparison with existing methods, the accuracy of product recommendations, and the accuracy of predicted review sentences, which were the issues in this experiment.

ACKNOWLEDGEMENTS

This research was supported by JSPS Grants-in-Aid for Scientific Research JP21H03496, JP22K12157, JP23H03688 and Sumitomo Electric Industries Group Social Contribution Fund.

In this paper, we used "Rakuten Dataset" (https://rit.rakuten.com/data_release/) provided by Rakuten Group, Inc. via IDR Dataset Service of National Institute of Informatics.

REFERENCES

Darius Afchar, Romain Hennequin. (2020). Making neural networks interpretable with attribution: application to implicit signals prediction. In *14th ACM Conference on Recommender Systems*, pp.220-229.

Riya Ambolkar, Arpita Bhagat, Bhakti Buga, and Swapnil Gharat. (2022). Hotel Recommendation System using advanced efficiency and accuracy with modified BERT technique. In *2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS)*

Dimple Chehal, Parul Gupta, and Payal Gulati. (2022). An Approach to Utilize E-commerce Product Reviews to Remove Irrelevant Recommendations. In *2022 IEEE Delhi Section Conference (DELCON)*

Taluya Hara, Jun Baba, and Takuya Iwamoto. (2021). Item-Driven Recommrndation: A Recommendation System for Other Items Based on the Users. In The *35th Annual Conference of the Japanese Socie 2021*

Taichi Imafuku, Tatsuya Kawakami, Tianxiang Yang, and Masayuki Goto. (2021). A Study on Item Recommendation Model by Evaluating the Effect of Individual Intervention. In The 35th Annual Conference of the Japanese Society for *Artificial Intelligence* 2021.

Rio Iwabuchi, Yoko Nakajima, Hirotoshi Honma, Haruka Aoshima, Akio Kobayashi, Tomoyoshi Akiba, and Shigeru Masuyama. (2017). Proposal of recommender system based on user evaluation and cosmetic ingredients. In 2019 *4th International Conference on Information Technology* (InCIT)

P.Kirubanantham, A.Saranya, and D.Senthil Kumar. (2022). Convolutional Recommended Neural Network system based on user reviews for movies.In *2021 4th International Conference on Computing and Communications Technologies* (ICCCT) .

Yuki Matsunami, Mayumi Ueda, Shinsuke Nakajima, Takeru Hashikami, Sunao Iwasaki, John O'Donovan, and Byungkyu kang. (2016). An automatic scoring method for review by evaluation item using a cosmetic item evaluation expression dictionary. In 8th *Forum on Data Engineering and Information Management* (DEIM Forum 2016) B1-1.

Hiromu Matsushima, Shun Morisawa, Takumi Ishikawa, and hayato Yamana. (2021). A Survey of Explainable

Recommender System. In 20th Information Science and Technology Forum (FIT2021).

Yoko Nakajima, Hirotoshi Honma, Haruka Aoshima, Akio Kobayashi, Tomoyoshi Akiba, and Shigeru Masuyama. (2019). Recommender System Based on User Evaluations and Cosmetic Ingredients. In Pakistan Journal of Engineering and Technology, PakJET Volume: 5, Number: 3, Pages: 38- 43.

Asami Okuda, Myumi Ueda, and Shinsuke Nakajima. (2020). A cosmetic product recommendation method using scores by evaluation item. In 12th Forum on Data Engineering and Information Management, B1-3.

Takayuki Onogawa, Ryohei Orihara, Yuichi Sei, Yasuyuki Tahara, and Akihiko Ohsuga. (2020). Why Do Users Choose a Hotel over Others? Review Analysis Using Interpretation Method of Machine Learning Models. In IEEE International Conference Big Data Analytics (ICBDA), pp.354-362.

Rakuten Group, Inc. (2014). Rakuten Dataset. Informatics Research Data Repository, National Institute of Informatics. (dataset). https://doi.org/10.32130/idr.2.0

Ashra Sahar, Muhammad Ayoub, Shabir Hussain, Yang Yu, and Akmal Khan. (2022). Transfer Learning-Based Framework for Sentiment Classification of Cosmetics Products Reviews. Pakistan Journal of Engineering and Technology, PakJET

Miharu Sakai, Mitsunori Matsushita, and Mayumi Ueda. (2019). Automatic Construction of Evaluation Expression Dictionary for Generating Scores of Cosmetics by Evaluation Item. In 11th Forum on Data Engineering and Information Management, B6-2.

Yuna Taniguchi, Asami Okuda, Mayumi Ueda, Panote Siriaraya, and Shinsuke Nakajima. (2019) Efficient method for creating evaluation expression dictionaries for each cosmetic category. In Proceedings of the 35th Symposium on Fuzzy Systems (FSS2019 Osaka University).

Sayaka Yabe, Mayumi Ueda, and Shinsuke Nakajima. (2021) Visualization system of differences among cosmetic items using scores by evaluation items, The 13th Forum on Data Engineering and Information Management, C14-2

Hafed Zarzour, Mohammad Alsmirat, and Yaser Jararweh. (2022). Using Deep Learning for Positive Reviews Prediction in Explainable Recommendation Systems. In 2022 13th International Conference on Information and Communication Systems (ICICS).