Feature Relevance Analysis of Product Reviews to Support Online Shopping

Fumiya Yamaguchi[1], Felix Dollack[1], Mayumi Ueda[2,3], and Shinsuke Nakajima[1]

[1] Kyoto Sangyo University, Motoyama, Kamigamo, Kita-ku, Kyoto 603-8555, Japan
{g1954390,felix,nakajima}@cc.kyoto-su.ac.jp
[2] University of Marketing and Distribution Sciences 3-1 Gakuen-Nishimachi, Nishi-ku, Kobe, Hyogo 651-2188, Japan
Mayumi_Ueda@red.umds.ac.jp
[3] Cybermedia Center, Osaka University, 1-32, Machikaneyama-cho, Toyonaka, Osaka, 560-0043, Japan

Abstract. The number of online shoppers has been increasing in recent years. Online shopping involves the risk that the purchased product may not be what was expected. Recently, the number of product review videos has also been increasing, and more users are using them as a reference because they provide a more accurate understanding of how the product is used than conventional reviews. With this development in mind, we have been developing a review video recommendation system to support online shopping. Our system helps users to know which product review videos they should watch. In this paper, we propose a review video feature analysis method, which is a necessary technology to realize the proposed system, and conduct two evaluation experiments to confirm the effectiveness of the proposed method. The results of the evaluation revealed that the proposed system received good ratings from the users, which confirmed the effectiveness of the proposed method.

Keywords: Comment Analysis · Product Reviews · Review Video · YouTube

1 Introduction

Online shoppers generally can only refer to images of product descriptions and reviews written by users who have purchased the products. If a users purchases a product based only on images of the product whose size cannot be compared or textual information in a review section, there is a risk that the product may not be what they were expecting. Therefore, we have recently seen an increase in the number of product review video postings. However, it is difficult to efficiently listen to only videos that provide the desired information because it is hard to know the contents of the videos in advance. Given this background, we believe that users will be able to watch videos more efficiently if they are informed about which feature of a product each video describes before playing the video. In this paper, we propose a feature analysis method to support online shopping and report the results of evaluation experiments.

2 Related Work

It is difficult to understand the contents of a video in advance, and it is difficult to efficiently listen to only those video with the desired information. Therefore, previous research has been conducted to classify videos based on the sentiment of video comments[1]. Similarly, Siersdorfer et al. propose a comment analysis method for YouTube using machine learning techniques[2]. In online shopping, reviews and word-of-mouth are important factors in making a decision to purchase a product. However, it is very time consuming to read through all review comments. Therefore, Haque et al. worked on sentiment of reviews using a machine learning method trained on an Amazon dataset[3]. Basani et al. summarized product reviews by classifying them into positive, negative, and neutral categories[4]. Zhang et al. performed sentiment analysis and opinion extraction of product reviews considering textual subjectivity[5]. Furthermore, Matsunami et al. constructed a dictionary of evaluation expressions based on review analysis that specialized in cosmetic items and developed an automatic review scoring method[6]. In addition, Scaffidi et al. proposed an automatic product retrieval that satisfies shoppers' requirements by extracting product features from reviews[7].

Thus, efforts aimed at summarizing the content of videos and reviews are flourishing. However, there have not been enough efforts to conduct feature relevance analysis of product review videos themselves. Against this backdrop, this study aims to conduct feature relevance analysis of product review videos for the purpose of online shopping support.

3 Scoring method by features

In recent years, YouTube[4] has become a well-known video-sharing site. A vast number of videos are uploaded to the above site on a daily basis, and product review videos are also actively published and viewed. When viewing product review videos during online shopping, it is important to know whether or not there are reviews of specific features among the multiple features (hereinafter referred to simply as "features") that serve as criteria for making a decision to purchase the product in question.

3.1 Processing steps of the feature analysis method for review videos

First, frequently appearing words in the video were extracted manually and classified to construct a dictionary for each feature of the product, as shown in the center of Fig.1. The higher the importance score, the more relevant each word is to each feature.

[4] YouTube,https://www.youtube.com/

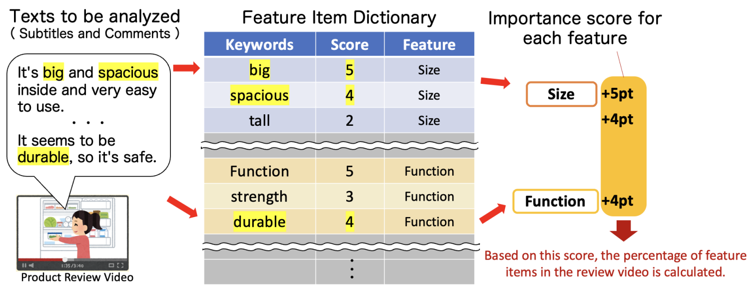

Fig. 1. Feature dictionary and automatic scoring of features in product review videos

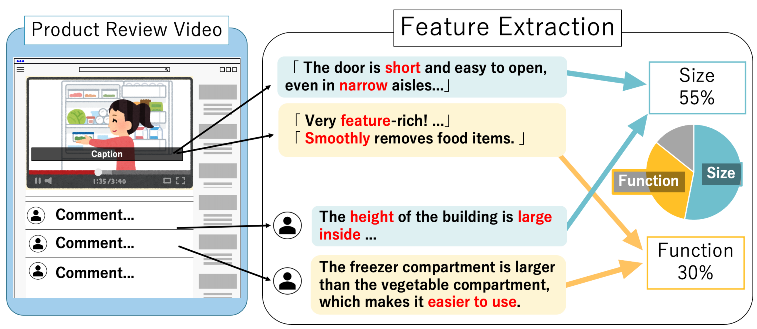

Fig. 2. Analysis method of feature relevance for product review videos

As for the construction of dictionaries, there are previous studies that worked on automatic dictionary creation. Taniguchi et al. have been working on improving dictionaries of evaluation expressions specialized for cosmetic items, and have studied efficient automatic score-rigging by constructing a dictionary of evaluation expressions based on Word2Vec[8]. However, it is difficult to fully automate the process from the viewpoint of accuracy.

Next, the presence or absence of feature expressions in the subtitle and comments attached to the video is checked using the constructed dictionary.

Finally, by totaling the number of sentences extracted for each feature, the ratio or importance of each product feature in the video is calculated. By performing the above processes, we aim to construct a system that enables the presentation of the ratio of features as shown in Fig.2.

4 Evaluation experiment

Two evaluation experiments were conducted. The first experiment was to evaluate the accuracy of the proposed method. The second experiment was to evaluate the usability of the proposed system. A total of 11 male participants in their 20s participated in both experiments.

We selected products that are relatively expensive as precision instruments and products that people are cautious about purchasing because they are not sure of the size or how to assemble them. The products selected were (1)Desk, (2)iPad, and (3)Camping Tent. Only videos that had been viewed more than 10,000 times were selected.

We limited the number of features to be extracted to five features per category. For review videos of desks and camping tents, we used the following five feature categories: Size, Function, Design, Cost performance, and Assembly. For review videos about the iPad, we use the following five features: Weight, Camera, Sound quality, Design, and Cost performance.

4.1 Accuracy evaluation of feature analysis method for video

Participants were asked to watch product review videos and respond to which extend they perceived the information of each feature was included in the videos. The average of the participants' answers is used as the Ground Truth data. Finally, we compare the feature proportions using Ground Truth and the feature proportions determined by the proposed method to verify the degree of agreement by Pearson correlation coefficient and cosine similarity.

Experimental results: In the (1)Desk video, the correlation coefficient and cosine similarity of the proposed method when both Ground Truth and subtitles and comments were analyzed were 0.98 and 0.98, respectively, and for subtitles only: 0.97 and 0.98. Thus, the cosine similarity for all other products showed high similarity. Furthermore, taking the average of the correlation coefficients for the three videos, we obtain 0.67 for the subtitle only and 0.47 for the subtitle and comment. Basically, the proposed method shows a strong correlation with Ground Truth, which confirms the effectiveness of the proposed method.

Consideration: The results of the accuracy evaluation experiments confirmed that the proposed method for feature relevance analysis of videos is generally effective. It is considered necessary to improve the accuracy of the proposed method by extending the dictionary or by other means. In addition, the number of videos used for evaluation was small because only one video in each category was used in this experiment. In the future, we would like to expand the number of videos used in the experiment and conduct a larger experiment that is less likely to be influenced by specific videos.

4.2 System Usability Evaluation

In the system usability evaluation experiment, the effectiveness of the proposed system was evaluated using the System Usability Scale (SUS) [9]. The SUS is a questionnaire that enables a numerical evaluation of system usability. The questionnaire consists of 10 questions with alternating positive and negative questions. The respondents were asked to answer on a 5-point scale from 1 (strongly disagree) to 5 (strongly agree). The results of the responses are transformed

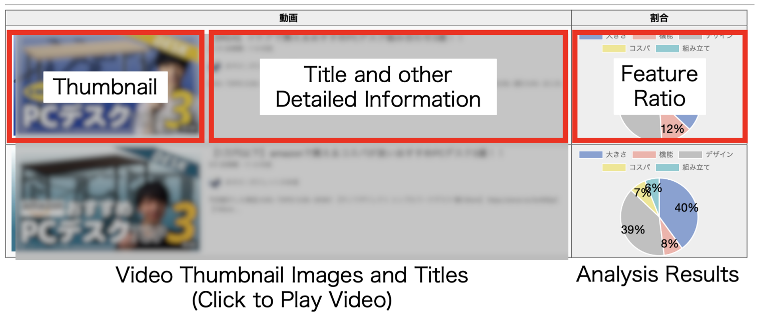

Fig. 3. Proposed system to be evaluated

and calculated on a 100-point scale to evaluate the usability of the system. The proposed system that was evaluated is shown in Fig. 3.

Table 1. System Usability Scale Questions

#	Question	Average Score (1 - 5)
1	I think that I would like to use this system frequently.	4.18
2	I found the system unnecessarily complex.	1.64
3	I thought the system was easy to use.	4.27
4	I think that I would need the support of a technical person to be able to use this system.	1.55
5	I found the various functions in this system were well integrated.	4.18
6	I thought there was too much inconsistency in this system.	1.73
7	I would imagine that most people would learn to use this system very quickly.	4.63
8	I found the system very cumbersome to use.	1.55
9	I felt very confident using the system.	4.45
10	I needed to learn a lot of things before I could get going with this system.	1.64

Experimental results and Consideration: The SUS question items and the average values of the participants' responses to each question are shown in Table 1. The results of the evaluation experiment show that the score of the proposed system is 84.09. The standard deviation between participants is 8.61. The table on the right side of Fig. 4 shows how to interpret the results of the SUS questionnaire. The proposed system was rated highly in all the questions. Considering the standard deviation, the proposed system is rated somewhere between A: Excellent and B: Good. The score of the proposed system is much higher than 68, which means that the proposed system has high system usability.

5 Conclusion

In this paper, we propose a feature relevance analysis method for product review videos to support online shopping. The feature relevance analysis of product review videos using a feature dictionary facilitates the judgment of which review videos could be of interest to the user and should be viewed. In the future we

SUS Score

Average	84.09
Standard Deviation	8.61

Grade Table

SUS score	Grade	Adjective Rating
> 80.3	A	Excellent
68 - 80.3	B	Good
68	C	Okay
51-68	D	Poor
< 51	E	Awful

Fig. 4. SUS Score of the proposed system

aim to improve the accuracy of the feature analysis method by expanding the feature dictionary. Based on the results of the evaluation experiments, we plan to add new features and improve the proposed method.

Acknowledgements This work was supported by multiple JSPS KAKENHI research grants (19K12243, 20H04293, 22K12281). It is also a product of research activity of the Institute of Advanced Technology and the Center for Sciences towards Symbiosis among Human, Machine and Data which was financially supported by a Kyoto Sangyo University Research Grant (M2001).

References

1. Abbi Nizar Muhammad et al. Sentiment Analysis of Positive and Negative of YouTube Comments Using Naïve Bayes – Support Vector Machine (NBSVM) Classifier, ICOMITEE: Int. Conf. on Computer Science, Information Technology, and Electrical Engineering, 2019.
2. Stefan Siersdorfer, Sergiu Chelaru, Jose San Pedro, Ismail Sengor Altingovde, and Wolfgang Nejdl. Analyzing and mining comments and comment ratings on the social Web. ACM Trans. Web 8, 3, Article 17 (June 2014), 2014.
3. Tanjim Ul Haque et al. Sentiment analysis on Large Scale Amazon Product Reviews. IEEE Int. Conf. on Innovative Research and Developmen, 2018.
4. Y. Basani et al. Application of Sentiment Analysis on Product Review E-Commerce. Int. Conf. on Advance and Scientific Innovation (ICASI), 1175 (2019).
5. Zhu Zhang et al. Utility Scoring of Product Reviews. The 15th ACM int. conf. on Information and knowledge management(CIKM'06), pp.51–57, 2006.
6. Y. Matsunami et al. Explaining Item Ratings in Cosmetic Product Reviews. IMECS 2016, 2016.
7. Christopher Scaffidi et al. Red Opal: Product-Feature Scoring from Reviews. EC '07: The 8th ACM conference on Electronic commerce,June 2007,pp. 182–191, 2007.
8. Mayumi Ueda et al. A Research on Constructing Evaluative Expression Dictionaries for Cosmetics Based on Word2Vec. iiWAS2021 pp.84-90, 2021.
9. J. Brooke, SUS—A Quick and Dirty Usability Scale, In P. W. Jordan, B. Thomas, B. A. Weerdmeester, A. L. McClelland, Eds., Usability Evaluation in Industry, Taylor and Francis, London, 1996.