A Product Feature Mentioned Timestamp Extraction Method in Review Videos for Online Shopping

Fumiya Yamaguchi
Kyoto Sangyo University
Kyoto, Japan
i2286211@cc.kyoto-su.ac.jp

Da Li
Fukuoka University
Fukuoka, Japan
lida@fukuoka-u.ac.jp

Mayumi Ueda
University of Marketing and Distribution Sciences
Hyogo, Japan
Mayumi_Ueda@red.umds.ac.jp

Shinsuke Nakajima
Kyoto Sangyo University
Kyoto, Japan
nakajima@cc.kyoto-su.ac.jp

***Abstract*—Online shopping entails the risk that the acquired product might not align with expectations. Consequently, an escalating number of users are turning to product review videos. Nonetheless, due to the profusion of evaluations proliferating across the internet, users encounter challenges when endeavoring to judiciously peruse the information germane to their interests. Henceforth, we are engrossed in the formulation of an analytical system designed to bolster the domain of online shopping. This system constitutes an analytical framework for the elucidation of product features, thereby expediting the discernment of which segments within a review video merit scrutiny. To elucidate, this system proffers the user with a quantification of product features delineated within the video, thereby affording streamlined video consumption. In this paper, we elucidate the development of a method tailored for the dissection of product features derived from the subtitles within review videos, concurrently ascertaining the "timestamps" of such mentions.**

***Index Terms*—Review Video, Product Reviews, YouTube, Subtitles**

I. Introduction

In recent years, there has been a surge in the number of online shoppers. However, online platforms, unlike physical stores, have the drawback of not allowing customers to physically inspect products. For instance, online shoppers typically rely solely on product images and user-generated reviews for their purchasing decisions. When a user buys a product solely based on images or textual descriptions in a review, without the ability to gauge its size or quality in person, there is a risk that the product may not meet their expectations. Consequently, there has been a notable rise in the posting of product review videos, which provide a more precise understanding of product usage and convey information that text or images alone cannot capture. As a result, an increasing number of users are turning to these videos as valuable references.

Efficiently identifying videos that offer desired information is challenging due to the sheer volume of review videos, many of which lack readily available content summaries. Against this backdrop, we posit that users could enhance their video-watching efficiency if they were informed about which aspects of a product each video covers prior to playback.

Building upon this premise, we proposed a dictionary-based product feature analysis system for review videos [1]. This endeavor has enhanced the efficiency of video consumption by quantifying feature mentions, thus simplifying the process of selecting videos that align with users' specific interests from a vast pool of review content. Nonetheless, even with the provision of feature ratios, watching each video in its entirety remains time-consuming. Moreover, pinpointing the desired scene within a video can be a time-intensive endeavor. Consequently, we posit that presenting mentioned timestamps is an effective means of expediting the acquisition of desired information, enabling users to achieve their objectives in a shorter timeframe (Fig. 1).

Recently, certain content creators on YouTube have started employing the chapter feature to facilitate a more comprehensible overview of their videos. Nonetheless, the chapters established by individual creators tend to be inconsistent, and a significant number of videos lack chapters altogether. Few videos offer a well-defined

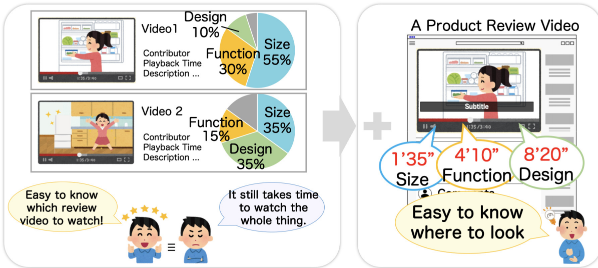

Fig. 1. Timestamps of mentioned product features in review video and issues with presenting only pie charts.

delineation of their content through chapter divisions. Additionally, the chapters introduced by creators do not align with the distinctive features of the products under review. The issue lies in the fact that these contributors predominantly clarify the video's structure rather than designating chapters based on the specific "product features" that we are interested in.

In this paper, we delineate a technique for discerning timestamps referencing the desired product features, which the user intends to explore. We conducted precision assessment experiments to substantiate the efficacy of the proposed methodology. Furthermore, we present findings from the validation of the review video analysis system we engineered, along with an assessment of its usability.

II. Related Work

Anticipating the content of a video in advance can be challenging. Consequently, prior research has sought to categorize videos based on the sentiment expressed in video comments. Notably, substantial efforts have been directed towards data mining and analytical approaches aimed at uncovering trends and significance by focusing on comments associated with YouTube videos. Ramamonjisoa et al. [2] classified YouTube comments into five distinct categories. Muhammad et al. [3] have concentrated on the classification of sentiments within comments. Furthermore, Khan et al. [4] prediced ambiguous data by scrutinizing comments on videos.

In the realm of online shopping, product reviews and word-of-mouth play pivotal roles in influencing purchasing decisions. Nonetheless, the arduous task of sifting through numerous review comments has prompted research into the automatic computation of ratings, leveraging factors such as content and contributors. Haque et al. [5] delved into sentiment analysis of reviews using machine learning techniques trained on an Amazon dataset. Furthermore, efforts have been made to visualize product reputation insights by dissecting product reviews. Furthermore, Matsunami et al. [6] and Ueda et al. [7] formulated a lexicon of evaluative terms, focusing on the analysis of reviews specific to cosmetic products. They subsequently devised an automated review scoring methodology.

All of the aforementioned studies have harnessed text analysis techniques. As previously noted, there is a burgeoning trend in employing Bidirectional Encoder Representations from Transformers (BERT) for text analysis. Agarwal et al. [8] delve into the application of sentiment analysis using BERT models for data originating from social media platforms. Similarly, Fimoza et al. [9] leverage BERT to conduct sentiment analysis on movie reviews posted on YouTube. While endeavors to summarize the content of videos and reviews are gaining momentum, there has been a paucity of efforts directed toward feature analysis of product review videos themselves. Furthermore, studies focusing on timestamp mentions have been limited.

In light of this context, our study embarks on a feature analysis of review videos with the aim of bolstering online shopping. More specifically, our emphasis is placed on timestamps within YouTube videos where product features are referenced.

III. Scoring Method Based on Features

In recent years, YouTube[1] has gained widespread recognition as a video-sharing platform. It sees a continuous influx of videos being uploaded daily, including a substantial volume of product review videos that are actively posted and viewed. When engaging with these product review videos during online shopping, it becomes pivotal to ascertain whether there are reviews pertaining to specific product attributes among the numerous criteria used to make informed purchase decisions (referred to hereafter simply as "features"). For instance, when contemplating the purchase of a refrigerator, the particular features of interest to a user, such as usability, size, operating noise, and functions, may vary.

As such, our objective is to furnish timestamps within each review video where these features are discussed. The process for identifying timestamps mentioning features in the review videos is delineated below.

A. *Scoring method based on features using dictionary*

1) Construction of feature-specific dictionaries: To facilitate the scoring method for product features, we

[1]https://www.youtube.com/

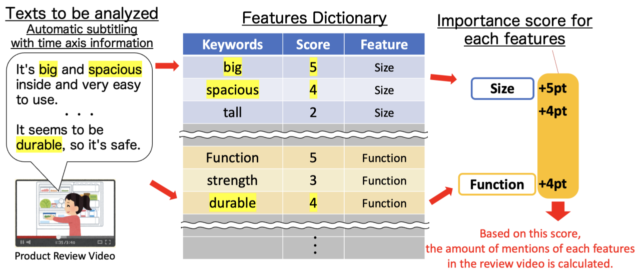

Fig. 2. Feature dictionary and automatic scoring of features in product review videos.

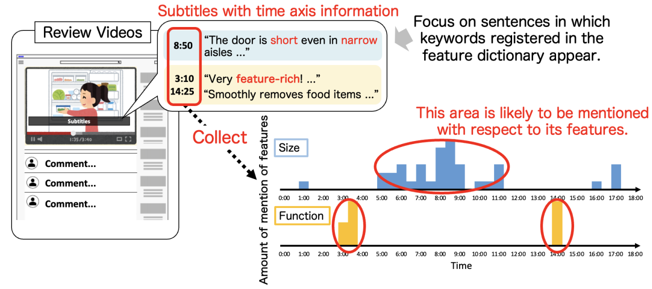

Fig. 3. Analysis method of feature items for product review videos.

initiated the process by assembling videos within each product category. Subsequently, we identified the features for categorization based on frequently occurring terms. By manually extracting commonly encountered terms within the videos, we categorized them to formulate individual dictionaries for each product feature, as depicted in the center of Fig.2. During the compilation of these dictionaries, words observed in the videos were collected and assigned significance scores on a per-word basis. The higher the assigned importance score, the more pertinent the word is considered to be in relation to each respective product feature.

Regarding the construction of dictionaries, prior research has explored the realm of automatic dictionary generation. Taniguchi et al. have dedicated their efforts to refining dictionaries of evaluative expressions tailored to cosmetic products. They have also explored efficient automated scoring methodologies by developing an evaluative expression lexicon grounded in Word2Vec [7]. Nevertheless, it is worth noting that achieving complete automation in this process remains challenging, particularly when considering accuracy as a critical factor.

2) Classification of subtitle text for each feature: Subsequently, we assess the presence or absence of feature-related expressions within the video's subtitles by employing the previously constructed dictionary. For every expression contained within a sentence, we consult the dictionary to ascertain the corresponding product feature it pertains to. By iteratively applying this process, all subtitle information and comment sentences are categorized under their respective product features (Fig. 2).

3) Extraction of mentions for each product feature in review videos: This section outlines the procedure for extracting mentions of each product feature within the review videos (Fig. 3). The analysis of mentions leverages automatic subtitles with time-stamped information, obtainable from the review videos. Subtitles that have been categorized under each feature also possess temporal information. Our focus lies in pinpointing the timing of appearances of keywords registered within the feature dictionary. These appearance times are aggregated for each feature and translated into a histogram, segmented into intervals of a few seconds. This approach enables us to decipher the specific scenes in which information about the products discussed in the video is presented. When product features are mentioned, peaks become evident in the histogram. This allows for an immediate understanding of where within the video to direct one's attention.

The resulting histogram generated by the system is not merely a simplistic tally of keyword occurrences. It draws upon insights from the study [1], which calculates the feature percentage within a video based on mention volume analysis. Consequently, the significance of keywords fluctuates in accordance with the scores registered in the dictionary, contributing to the creation of a smooth histogram. Through the aforementioned process, our objective is to establish a system capable of presenting the number of mentions for each timestamp categorized by feature.

B. System Implementation for displaying mentioned timestamps of product features

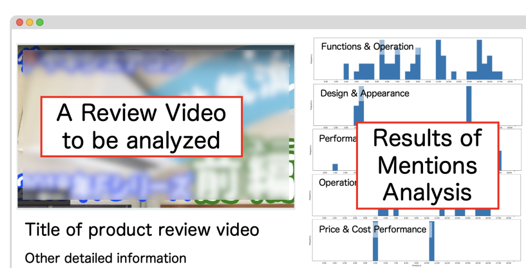

Fig. 4. Conceptual diagram of our system.

Here is an illustration of the system under development (Fig. 4). On the left, users can observe a YouTube

TABLE I
THE FEATURE ITEMS.

Features	Information about the products included in the review videos
Function	Durability, strength, additional functions
Design	Color, appearance texture
Performance	Ability and how much power it has
Noise	Operating noise, vibration of outdoor units
Cost performance	Price to Value

review video, and on the right, there is a graph depicting mentions related to the product. This graph represents a chronological histogram of product feature mentions. If users are interested in a particular product, they can refer to this histogram. It serves as a valuable tool for pinpointing timestamps to watch, enabling users to efficiently gather information about the product.

C. Setting of target products and features

We opted to focus on relatively expensive appliances and products that often prompt caution among potential buyers due to uncertainties related to size or user experience. Our target products encompass home appliances such as “air conditioners” and “refrigerators.” To streamline the process, we limited the number of features to be extracted to five key attributes. These five features were broadly categorized as follows: function, design, performance, noise, and cost performance (Table I).

The determination of these features was based on the analysis of subtitles and comments within videos belonging to each product category. We compiled a list of selected terms, ranked by their frequency of appearance, and employed them as the basis for evaluating these features. The proposed method extracts information pertaining to the relevance of these features within product review videos.

IV. EVALUATION EXPERIMENT

In this section, we delineate the evaluation experiments conducted to assess the proposed method presented in Section III. Four distinct evaluation experiments were undertaken to gauge the effectiveness of the proposed method. The first experiment entails a comparison of the proposed method with a conventional approach, while the second experiment involves a comparison with a baseline method. The third experiment assesses the usability of the proposed system, and the fourth examines the system’s usefulness and comprehensibility.

To commence, we expound upon the comparison experiments involving the proposed method and conventional methods. A total of 8 male participants in their 20s took part in both experiments. The review videos for evaluation were sourced from YouTube.

A. Comparative experiments with conventional methods

*a) **Outline of the Experiment**:* To assess the effectiveness of the proposed method, a comparative analysis was conducted between the proposed approach and the conventional method. The proposed method involves viewing the review video with the presentation of analyzed the mentioned timestamp content, whereas the conventional method entails watching the review videos without the benefit of timestamp presentation. The comparison was made in terms of the time and complexity required to gather information using each method. The specific experimental procedure is outlined as follows:

1) Present a quiz to the participants regarding how the product is reviewed in the video.
2) Participants are asked to find the information indicated in the review video.
3) Measure the time it takes participants to discover the answer to a question.
4) Participants respond with a subjective rating (5-point scale) of annoyance for the above initiatives.

This process was carried out for both the conventional method and the proposed method. Comparisons were made based on the “time” and “subjective evaluation” obtained through the aforementioned procedure. “Subjective evaluation” serves as an indicator for assessing the level of annoyance experienced while searching for information in a video. Participants were asked to respond to the following three questions on a five-point scale (ranging from 1, “Not applicable at all,” to 5, “Very applicable”).

- I found the answer right away.
- I was able to watch only the relevant parts.
- I want to do this frequently.

Concerning the “time” comparison, the effectiveness of the proposed method is confirmed when the time required for the proposed method is shorter than that of the conventional method. Regarding the “subjective evaluation (5 levels),” the effectiveness of the proposed method is established when the score for the proposed method is higher than that for the conventional method.

*b) **Experimental results**:* To begin, the time required to locate the specified information within the review videos is presented in Tables II and III. These tables display the time in seconds (s) needed for the conventional method and the proposed method. The

TABLE II
COMPARISON RESULTS OF THE TIME TAKEN TO FIND THE INDICATED INFORMATION FOR EACH VIDEO.

Video #	Conventional method	Proposed Method
1	169.50	237.00
2	125.00	141.00
3	158.00	188.75
4	183.75	118.00
5	143.75	118.00
6	154.75	60.00
7	146.50	145.25
8	162.00	64.25
Average	155.41[sec]	134.03[sec]

TABLE III
COMPARISON RESULTS OF THE TIME TAKEN TO FIND THE INDICATED INFORMATION FOR EACH PARTICIPANT.

Participant #	Conventional method	Proposed Method
1	105.25	73.50
2	149.75	222.75
3	138.75	105.50
4	141.75	247.25
5	172.25	71.00
6	106.25	95.25
7	201.50	133.25
8	227.75	123.75
Average	155.41[sec]	134.03[sec]

time taken for each video and each participant was calculated. Additionally, a graph illustrating all the time data is provided for a comparative analysis between the proposed and conventional methods (see Fig. 5).

Tables II and III reveal that the average time per participant for one video was 155.4 seconds for the conventional method and 134.0 seconds for the proposed method. In 5 out of 8 videos, the proposed method facilitated the discovery of necessary information in less time compared to the conventional method, which often required more time. Among the 8 participants, the proposed method was more time-efficient for 6 of them.

Table IV shows that the proposed method outperforms the conventional method in all evaluation items. We performed significance level 0.1% (0.001) t-tests on all questions and confirmed that there was a difference in dominance.

*c) **Consideration**:* The results presented above strongly demonstrate the effectiveness of the proposed method in all aspects of the five-level subjective evaluation. Moreover, the actual time taken by the proposed method is shorter than that of the conventional method, indicating a reduced time requirement for finding the desired information. The overall effectiveness of the proposed method is visually evident from the comprehensive data graph (Fig. 5). While Table II indicates that the first three outcomes for the proposed method may appear suboptimal, it is worth noting that the proposed method exhibits effectiveness in the latter five cases.

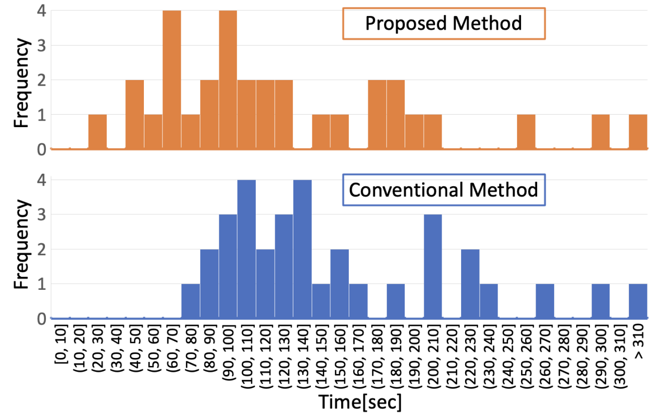

Fig. 5. Comparison with histograms of all times.

Several factors may have contributed to the relatively small differences observed in the time comparisons between the proposed and conventional methods. (1) Some of the analysis results may not have been entirely appropriate. (2) Participants might have spent additional time comprehending the experiment's structure and how to interpret the "mention analysis graph." (3) The proposed method involved responding to several more questions than the conventional method, which naturally extended the time required. However, since the same questions were repeated, participants likely became more efficient in responding as the survey progressed. In summary, we believe that the proposed method enhances the likelihood of acquiring sensory information, favoring video content over reviewing text or image-based information alone, as typically found in reviews.

B. *Evaluation of usefulness and clarity*

*a) **Outline of the Experiment**:* Following the completion of the experiment comparing the conventional and proposed methods as described earlier, participants were queried about the utility and comprehensibility of the proposed method. The questionnaire encompassed the following two inquiries:

- I think the graph showing where the features are mentioned was helpful.
- The graphs showing where the features were mentioned were easy to understand.

These questions assess the system's utility and comprehensibility, respectively. The rating system employed a 5-point scale, ranging from (1) "Not at all" to (5) "Very much." In essence, the system's effectiveness could be

TABLE IV
COMPARISON OF 5-LEVEL SUBJECTIVE RATINGS.

Questions	Conventional method	Proposed Method
I found the answer right away.	3.44	4.56
I was able to watch only the relevant parts.	2.72	4.34
I want to do this frequently.	2.59	4.53
Average	2.92	4.48

TABLE V
RESULTS OF USEFULNESS AND CLARITY EVALUATION FOR THE PROPOSED METHOD.

Assessment Items	Questions	Score
Usefulness	I think the graph showing where the features are mentioned was helpful.	4.69
Comprehensibility	The graphs showing where the features were mentioned were easy to understand.	4.47

validated if the median value surpassed (3) “neither agree nor disagree.”

*b) **Experimental results and Consideration**:* The results are presented in Table V. The evaluation results indicated that the mean of the participants’ responses exceeded the median. In the free-text comments section, several positive remarks were made, such as “I believe it is highly effective for longer videos” and “I consider it beneficial for various types of videos.” However, some participants suggested room for improvement, particularly regarding minor discrepancies between the analyzed and indicated segments and the visibility of the graph display.

V. CONCLUSION

In this paper, we proposed a method for analyzing timestamps in review videos where product features are discussed, with the aim of supporting online shopping. Our method enhances the efficiency of watching lengthy product review videos by pinpointing where specific product features are mentioned, ultimately reducing the risk associated with online shopping. The accuracy of the proposed method and the effectiveness of the associated system were validated through experiments. Looking ahead, future work will involve the utilization of machine learning for the analysis of review videos and the conduct of additional evaluation experiments.

ACKNOWLEDGMENT

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 (M2301).

REFERENCES

[1] F. Yamaguchi, F. Dollack, M. Ueda, and S. Nakajima, “Feature relevance analysis of product reviews to support online shopping,” International Conference on Information Integration and Web Intelligence (iiWAS2022), 2022, pp. 441-446.

[2] D. Ramamonjisoa, H. Ikuma, and R. Murakami, “Filtering relevant comments in social media using deep learning,” 2022 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), 2022, pp. 335-340.

[3] A. N. Muhammad, S. Bukhori, and P. Pandunata, “Sentiment analysis of positive and negative of YouTube comments using naïve bayes – support vector machine (NBSVM) classifier,” 2019 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE), 2019, pp. 199-205.

[4] F. E. Khan, A. M. Ruhan, R. Shamsuddin, and F. B. Ashraf, “A machine learning approach to predict movie success from Youtube trailer comments,” 25th International Conference on Computer and Information Technology (ICCIT), 2022, pp. 372-377.

[5] T. U. Haque, N. N. Saber, and F. M. Shah, “Sentiment analysis on large scale Amazon product reviews,” IEEE International Conference on Innovative Research and Development (ICIRD), 2018.

[6] Y. Matsunami, M. Ueda, and S. Nakajima, “Explaining item ratings in cosmetic product reviews,” International MultiConference of Engineers and Computer Scientists (IMECS), 2016.

[7] M. Ueda, Y. Taniguchi, D. Li, P. Siriaraya, and S. Nakajima, “A research on constructing evaluative expression dictionaries for cosmetics based on Word2Vec,” International Conference on Information Integration and Web Intelligence (iiWAS2021), 2021, pp. 80-86.

[8] M. Agarwal, P. K. Chaudhary, S. K. Singh, and C. Vij, “Sentiment analysis dashboard for socia media comments using BERT,” 2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT), 2023, pp. 292-298.

[9] D. Fimoza, A. Amalia, and T. H. F. Harumy, “Sentiment analysis for movie review in Bahasa Indonesia using BERT,” 2021 International Conference on Data Science, Artificial Intelligence, and Business Analytics (DATABIA), 2021, pp. 27-34.