An Impression-Based Video Analysis Method for Video Recommender Systems

Akito Sasahara
Kyoto Sangyo University
Kyoto, Japan
g2053507@cc.kyoto-su.ac.jp

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 Science
Hyogo, Japan
Mayumi_Ueda@red.umds.ac.jp

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

***Abstract*—In recent years, there has been a significant increase in the amount of video content available on video sharing platforms. Users often rely on titles and thumbnails to make their video selections based on their interests. However, it is challenging to determine whether a video truly meets a user's expectations until it is played. To address this issue, this study focuses on video impressions, analyzing comments posted on videos and assigning intuitive evaluation scores for various aspects of the video. By presenting these evaluation scores before watching a video, it becomes possible to recommend videos from a fresh perspective based on the user's intuition. This approach empowers users to assess whether a video aligns with their preferences before committing to play it. We have developed an intuitive video recommender system based on this concept. In this paper, we elaborate on the impression-based video analysis method and present the results of evaluation experiments conducted using the developed intuitive video recommender system.**

***Index Terms*—Video Impressions Analysis, Intuitive video retrieval, Automatic Scoring, Video Recommendation**

I. Introduction

In today's digital landscape, the abundance of video content on video sharing platforms has resulted in users heavily relying on titles and thumbnails to make their video selections [1]. However, traditional keyword searches have limitations when it comes to providing intuitive video recommendations that cater to users' preferences. For instance, conducting a keyword search for "Emo" may yield numerous videos containing the keyword in their title or description (as depicted in Table I). However, there are many videos embodying the essence of "Emo" even if the keyword itself is not explicitly mentioned in the title. Therefore, the current simplistic approach of keyword searches falls short in enabling intuitive video recommendation.

In this study, we propose an innovative approach that facilitates impression-based video analysis through the use of an automatic comment scoring method for each evaluation item. The primary objective is to develop a system that allows users to discover videos they desire to watch using intuitive keywords, such as "Emo," which is commonly used in Japan to refer to videos with an emotional or sentimental atmosphere. By meticulously analyzing the comments posted on videos, our method assigns a score to each evaluation item based on users' intuitive assessments of the video's qualities. By presenting these scores before a user starts watching a video, we enable them to explore videos from a fresh perspective guided by their intuition. This approach empowers users to determine whether a video truly aligns with their preferences before engaging with it. In this paper, we provide a comprehensive outline of the proposed system and present the results of an evaluation experiment that validates the system's effectiveness.

II. Related Work

In recent years, popular video sharing platforms such as YouTube[1] and Niconico[2] have allowed users to freely comment on videos, resulting in a massive influx of comments due to the vast number of daily video uploads.

[1]https://www.youtube.com/
[2]https://www.nicovideo.jp/

Users rely on tag searches or keyword searches to find videos of interest and can provide their evaluations and comments on these videos.

In a related study focusing on video analysis, Murakami et al. proposed a method to quantify the level of "laughter" in videos by analyzing the presence of the character "w," which is commonly used in Japan to represent laughter [2]. Nakamura et al. developed a regular expression dictionary based on sensibility and utilized it to analyze user comments, enabling ranking searches based on sentiments such as positive, negative, happy, sad, and surprise [3]. Hanif Bhuiyan et al. introduced a natural language processing (NLP)-based sentiment analysis for user comments on YouTube videos to enhance the ranking of search results [4].

In this study, we aim to go beyond the existing sentiment analysis approaches and develop a technology that enables users to perform more intuitive video retrieval. While previous studies have focused on automatic scoring methods based on review analysis [5], Matsunami et al. developed an automatic review scoring method by creating a specialized dictionary of evaluation expressions for cosmetic items [6]. Building on these dictionary construction methods, this study seeks to achieve a more intuitive selection of evaluation items for videos and an efficient extraction of evaluation expressions for each item. Although there have been numerous studies on video recommendation through video comment analysis and scoring methods for individual evaluation expressions based on review analysis [7], there has been no prior research and development on an automatic scoring method for each evaluation expression specifically tailored to intuitive video recommendation, which is the primary objective of this study.

III. Automatic Comment Scoring System by Evaluation Item

A. Overview of the Proposed Method

We present a method that analyzes comments associated with videos and applies automated scoring based on evaluation items. This approach enables users to determine whether a video aligns with their expectations before playing it. As depicted in Figure 1, conventional methods rely on users selecting videos from a list on video sharing sites primarily based on titles and thumbnails, which may not accurately reflect the content they anticipate. In contrast, our proposed method assigns scores to evaluation items specifically tailored to videos, surpassing the limitations of conventional keyword searches and simple sentiment-based searches, such as positive or negative. This empowers users to assess whether a video genuinely meets their expectations even before watching it.

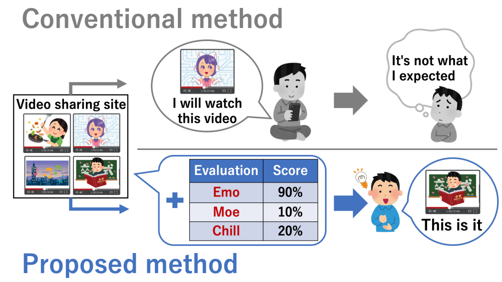

Fig. 1. Overview of the proposed method.

B. Consideration of Evaluation Items

In this section, we discuss the evaluation items used for video recommendations. The objective of this study is to provide intuitive video recommendations that go beyond simple sentiment-based searches, such as positive or negative. The evaluation items considered in this paper encompass nine aspects: "Emo," "Chill," "Kyun," "Jiwaru," "Pien," "Iratsuku," "Guro," "Bae," and "Moe."

The meanings of the currently proposed evaluation items are provided in Table 1.

TABLE I
Meaning of Evaluation Items.

"Emo"	It signifies the evocation of a nostalgic and emotional feeling.
"Chill"	It signifies the embodiment of a relaxed and leisurely vibe derived from the phrase "chill out."
"Kyun"	It signifies the expression of adorableness and butterflies in the stomach.
"Jiwaru"	It signifies the conveyance of a slowly building sense of comedic amusement.
"Pien"	It signifies the evocation of emotions that can be described as both "sadness" and "inspiration," evoking tears.
"Irastuku"	It signifies a state of being angered or irritated
"Guro"	It signifies the representation of something gruesome and highly unsettling, derived from the word "grotesque."
"Bae"	It signifies the representation of things that are vivid, beautiful, and possess a photogenic quality.
"Moe"	It signifies the emotion of "adorable" primarily towards anime characters.

C. Construction of Classifier Models for Each Evaluation Item

In this section, we outline the process of constructing a classifier model for each evaluation item, accompanied by corresponding scores. Users often express their intuitive evaluations of videos through a wide range of expressions in the comments they post. To capture this diversity, we collect actual intuitive evaluation expressions used by users and utilize them as training data to construct individual classifier models for each evaluation item. The training data for these classifier models are sourced from comment data obtained through the YouTube Data API[3], provided by the popular video-sharing website YouTube. We employ fastText, a library for efficient text classification, to construct the classifier models. To illustrate the procedure, we present an example of building a Kyun classifier model.

Kyun Classifier Construction Procedure

1) Collect a set of videos that represent the Kyun evaluation item, and gather 5,000 comments associated with each video.
2) Utilize fastText to create a multi-label classification model for all nine evaluation items.
3) To gather non-Kyun comments, apply the constructed classifier from step 2 to score 200 videos.
4) Identify videos with an Kyun score of 0.05 or lower as non-Kyun videos, and collect 5,000 comments from this set.
5) Train the Kyun classifier model using a combination of 5,000 Kyun comments and 5,000 non-Kyun comments, employing fastText.

Following a similar procedure, classifier models are constructed for the remaining eight evaluation items. As depicted in Figure 2, when comments from unknown videos are collectively evaluated through each classifier model, individual scores are generated for each evaluation item.

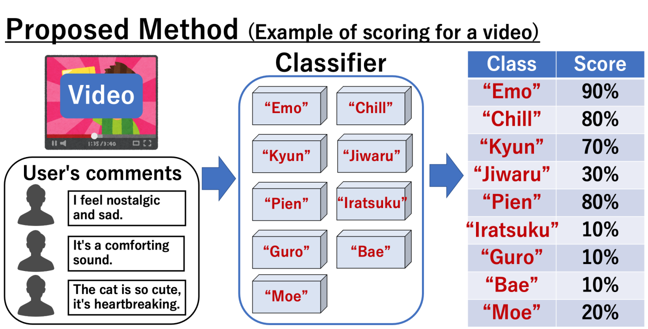

Fig. 2. Procedure of the proposed method.

[3]https://developers.google.com/youtube/v3

IV. Development of an Intuitive Video Recommender System

A. Video Data Used in the Intuitive Video Recommender System

The system incorporates a dataset of 700 videos. A detailed breakdown of the videos is provided below:

- The top 10 videos are selected in order of relevance for each evaluation item (e.g., "Emo") by conducting a retrieval on YouTube. This results in a total of 90 videos (10 videos per evaluation item).
- An additional set of 610 videos is included, which are manually curated based on their relevance to the evaluation items.

B. How to Use the Intuitive Video Recommender System

To utilize the intuitive video recommender system, users should follow the instructions provided below:

- In the interface, locate the section labeled "Select the evaluation item to be searched".
- Choose the desired evaluation item from the available options and move it to the "Priority of the evaluation item" section.
- Initiate the retrieval, which will generate the recommendation result screen (see Figure 3).
- The videos will be displayed in order of their "Kyun" score, accompanied by a radar chart corresponding to each score. It enables easy video comparison and provides insights into the video's characteristics even before watching it.
- The video recommender system also allows users to search using multiple evaluation items or a combination of keywords and evaluation items.

By following these steps, users can effectively explore and find videos based on their desired evaluation items, facilitating a more intuitive video retrieval experience.

V. Evaluation Experiment

A. Ranking Performance Comparison Experiment

In this section, we describe an experiment designed to compare the ranking performance of the proposed method with that of the conventional method.

1) Experiment Participants: The experiment involved a total of 19 participants, consisting of 18 males and 1 female, all in their 20s.

2) Experiment Procedures: To compare the ranking performance between the conventional method (YouTube) and the proposed method (intuitive video recommender system), we employed an interleaving method

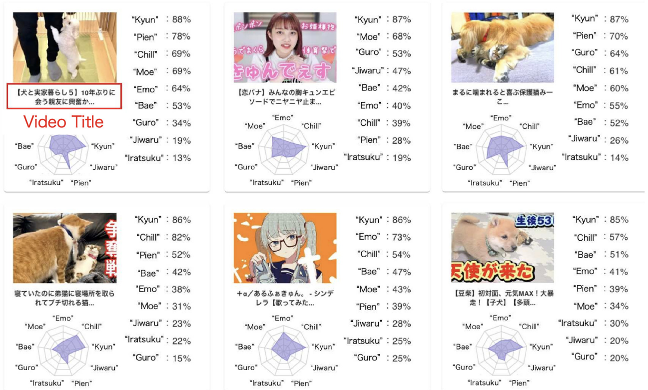

Fig. 3. Examples of the video recommendation results.

for the evaluation. The interleaving method used in this experiment was based on the work of [8] [9] [10].

As an example, we will describe the evaluation method for the “Kyun” ranking. As depicted in Figure 4, a keyword search for the term “Kyun” is performed using the conventional method, and the top 10 videos in order of relevance are considered as the top 10 “Kyun” ranking for the conventional method. Similarly, the proposed method retrieves for “Kyun” and ranks the top 10 videos based on their “Kyun” scores as the top 10 videos for recommendation.

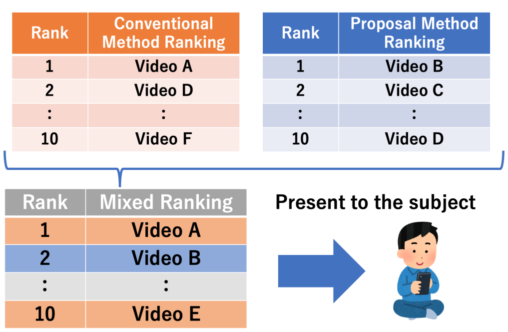

Fig. 4. Overview of the interleaving experiment.

The video set used in this evaluation is described in Section IV. A mixed ranking of “Kyun” is generated from these two rankings. The generation of the mixed ranking follows the Probabilistic Interleaving method [11]. As depicted in Figure 4, the generated top 10 “Kyun” mixed ranking is presented to the participants. They are then asked to intuitively click on the videos that they perceive as “Kyun” and watch them. As shown in Figure 5, the top 10 “Kyun” mixed ranking is presented to the participants. They are then instructed to click and

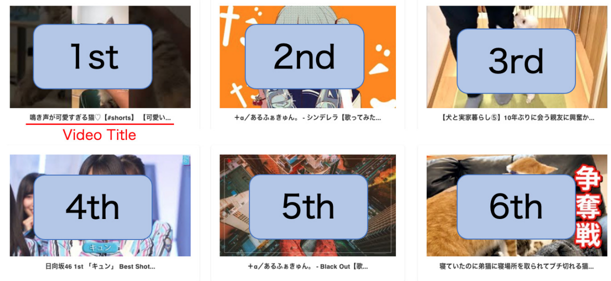

Fig. 5. The mixed ranking of “Kyun”.

watch the videos they intuitively find as “Kyun”.

During the video viewing process, participants are provided only with the thumbnail and title of the videos. As a precautionary measure, participants are asked to watch the videos in their entirety. However, participants have the option to stop watching a video if they perceive it to be “Kyun”, even if it is a long video such as background music for work, skipping 5 to 10 minutes of content. After watching each video, participants are requested to provide a 5-point rating indicating whether they felt it was “Kyun” or not (see Figure 6). If their response is “Yes” or “Slightly Yes,” they are given credit. This process is repeated until there are no more videos that are intuitively perceived as “Kyun”.

The same procedure is then repeated for the remaining evaluation items, including “Chill”, “Kyun”, “Jiwaru”, “Pien”, “Iratsuku”, “Bae”, and “Moe”. Due to ethical concerns, an experiment for the “grotesque” evaluation item was not conducted.

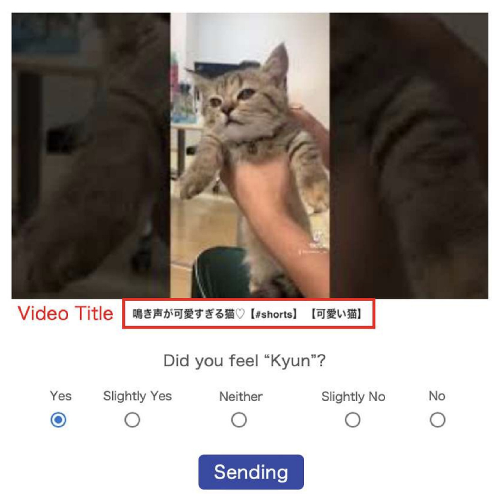

Fig. 6. Screen for identifying evaluation items after video viewing.

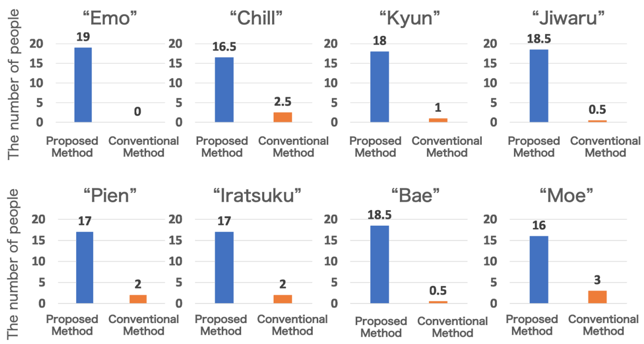

Fig. 7. Probabilistic interleaving results of each class.

3) Experiment Results: Figure 7 presents the results of the Probabilistic Interleaving for each evaluation item. In the event of a tie, 0.5 participants are assigned to both the conventional method and the proposed method.

For the “Emo” ranking, the proposed method had 19 participants, while the conventional method had 0 participants. This indicates that the proposed method outperformed the conventional method in terms of the “Emo” ranking.

In the “Chill” ranking, the proposed method had 18 participants, and the conventional method had 1 participant. The proposed method demonstrated superiority over the conventional method for the “Chill” ranking.

Similarly, for the “Kyun,” “Jiwaru,” “Pien,” “Irastuku,” “Bae,” and “Moe” rankings, the proposed method outperformed the conventional method, with higher participant counts favoring the proposed method in each case.

Table II displays the p-values obtained from the t-test (Correspondence, two-tailed, 5% level of significance) for each evaluation item. The presence of significant differences across all evaluation items confirms that the performance of the proposed method differs significantly from that of the conventional method.

Overall, the experiment results demonstrate the superiority of the proposed method across all evaluation items.

4) Consideration: The results of Probabilistic Interleaving demonstrated that the proposed method outperformed the conventional method in terms of ranking performance for all evaluation items. Moreover, significant differences were observed across all evaluation items. These findings suggest that there are videos that are considered better than those obtained solely through conventional keyword searches. It was evident that videos with intuitive evaluation comments related to the evaluation item names are more likely to align with users’ true preferences, compared to videos that merely include the evaluation item names in their title or description.

B. Intuitive Video Recommender System Evaluation Experiment

In this section, we outline an evaluation experiment.

1) Experiment Participants: The evaluation experiment involved 14 male participants, all in their 20s.

2) Experiment Procedure: We conduct an experiment to evaluate the intuitive video recommender system, focusing on the “Kyun” video recommendation as an example.

1) Participants were instructed to perform a keyword search for “Kyun” using the conventional method (YouTube) and select one video they intuitively want to watch.
2) Using the proposed method (intuitive video recommender system) described in Section 4, participants were asked to search for “Kyun” and select one video they want to watch intuitively based on the thumbnail, title, score, and radar chart provided by the system.
3) Participants were then complete the original questionnaire presented in Table III. The questionnaire is designed to compare and evaluate the proposed system against the conventional system. Participants can choose from five options: “The proposed system is better,” “The proposed system is somewhat better,” “The proposed system is the same as the conventional system,” “The conventional system is somewhat better,” and “The conventional system is better.” The corresponding scores are 5, 4, 3, 2, and 1, respectively.

The above procedure was repeated for each evaluation item. Participants were divided into two groups: Group 1 (conventional method followed by proposed method) and Group 2 (proposed method followed by conventional method). After completing the original questionnaires for all evaluation items, participants were asked to complete the System Usability Scale (SUS) questionnaire for both the conventional and proposed methods. The questionnaire items are shown in Table IV.

Finally, participants were asked to provide feedback in free text regarding what they liked about the intuitive video recommender system and any suggestions for improvement.

3) Experimental Results and Discussion: The results of the comparison with the conventional method, based on the original questionnaire, are presented in Figure 8,

TABLE II
PROBABILISTIC INTERLEAVING P-VALUE OF EACH EVALUATION ITEM.

	Emo	Chill	Kyun	Jiwaru	Pien	Iratsuku	Bae	Moe
p-value	4.33×10^{-28}	1.45×10^{-3}	6.64×10^{-7}	5.89×10^{-13}	4.74×10^{-6}	4.74×10^{-6}	5.89×10^{-13}	7.15×10^{-5}

TABLE III
QUESTIONNAIRE FOR EVALUATION EXPERIMENT.

A	Easy to find the video you want to watch.
B	Easy to compare video ratings before viewing.
C	Little gap in impressions of the video before and after viewing.
D	Easy to understand evaluation for each video.
E	There were many videos of interest.

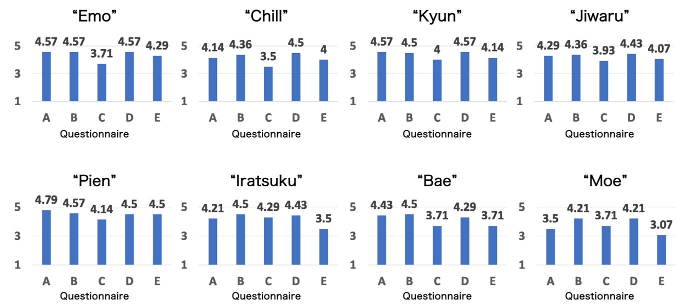

Fig. 8. Results of our original questionnaire.

TABLE IV
SYSTEM USABILITY SCALE QUESTION ITEMS.

1	This system allows intuitive searching.
2	I found the system unnecessarily complex.
3	This system minimizes the gap between impressions of the video before and after viewing.
4	I think that I would need the support of a technical person to be able to use this system.
5	I think that I would like to use this system frequently.
6	I found the system very cumbersome to use.
7	I thought the system was easy to use.
8	I thought there was too much inconsistency in this system.
9	I felt very confident using the system.
10	I needed to learn a lot of things before I could get going with this system.

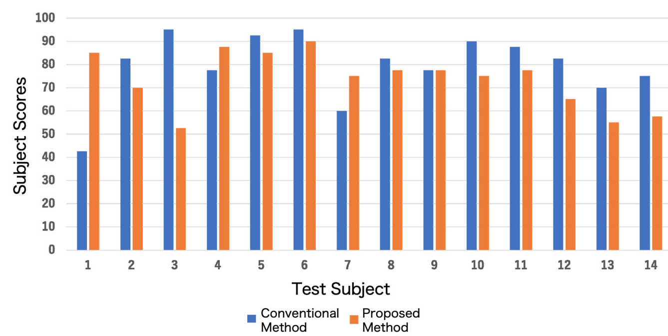

Fig. 9. SUS results of each Test Subject.

while the results based on the SUS for each subject are shown in Figure 9.

Upon analyzing the results from our original questionnaire, as displayed in Figure 8, it is evident that a significant number of participants rated the proposed system as better overall, indicating its higher level of intuitiveness compared to the conventional system.

Examining Figure 8, which illustrates the responses for "Little gap in impressions of the video before and after viewing," it can be observed that the average scores for "Emo," "Chill," "Bae," and "Moe" were in the 3-point range, indicating similarity to the conventional system. This implies that many videos aligned with the expectations set by their thumbnails and titles. Conversely, the average scores for "Kyun," "Jiwaru," "Pien," and "Iratsuku" were in the 4-point range or close to it, suggesting an improvement with the proposed system. Furthermore, the presentation of scores for each evaluation item helped to reduce the gap. The reason for the discrepancy in ratings for each evaluation item lies in the fact that numerous videos maintained consistency with their thumbnails throughout, resulting in minimal change in impressions. On the other hand, the "Kyun," "Jiwaru," "Pien," and "Iratsuku" categories encompassed videos where impressions changed during the course of the video, beyond the impact of the thumbnail and title.

Next, we turn our attention to the SUS results presented in Figure 9. It is worth noting that only three participants rated the conventional method higher, whereas the average score for the conventional method in the SUS was 73.58, while the proposed method obtained an average score of 79.29. This signifies that the usability of the proposed method surpasses that of the conventional method. However, based on a t-test analysis (two-tailed, correspondence, at a significance level of 5%), no significant difference was observed between the two methods. Focusing on the aspect "This system minimizes the gap between impressions of the video before and after viewing." in Figure 8, the average score for the conventional method was 3.14, whereas the proposed method garnered an average score of 4.21. Furthermore, a t-test (two-tailed, with correspondence, at a significance level of 5%) revealed a significant difference. In summary, the intuitive video recommender system enables users to ascertain whether a video truly aligns with their expectations even before watching it, in contrast to conventional systems. Finally, we summarize the positive aspects and areas for improvement of the proposed system based on the provided free-text descriptions as follows:

Positive aspects

- The radar chart facilitates easy comparison and identification of videos.
- The actual videos closely matched the expectations created by the thumbnails and titles.
- The system presented a larger number of videos that corresponded to the intended emotions compared to the conventional method.
- The recommendation results showed reduced bias compared to conventional methods.

Areas for improvement

- Enhancements to the user interface
- Implementation of more flexible recommendation methods to accommodate user preferences
- Consideration of variations in users' interpretations and impressions of each evaluation item

VI. Conclusion

The primary objective of this paper was to develop an intuitive video recommender system that allows users to retrieve videos based on their intuition and ascertain whether a video meets their expectations before watching it. Through the proposed system, we conducted evaluation experiments, which demonstrated its superiority over conventional methods in terms of ranking performance for all evaluation items. Moreover, the introduction of evaluation scores bridged the gap between previewing expectations and post-viewing impressions.

Future research could focus on further refining the system's accuracy and expanding the range of evaluation item categories, thus providing an even more comprehensive and intuitive video search experience.

Acknowledgment

This work was supported by multiple JSPS KAKENHI research grants (19K12243, 20H04293, 22K12281), and the Institute of Advanced Technology and the Center for Sciences towards Symbiosis among Human, Machine and Data which was supported by a Kyoto Sangyo University Research Grant (M2301).

References

[1] X. Wu, C.-W. Ngo, A. G. Hauptmann, and H.-K. Tan, "Real-time near-duplicate elimination for web video search with content and context," *IEEE Transactions on Multimedia*, vol. 11, no. 2, pp. 196–207, 2009.

[2] N. Murakami and E. Ito, "Emotional video ranking based on user comments," in *iiWAS '11: Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services*, 2011, pp. 499–502.

[3] S. Nakamura and K. Tanaka, "Video search by impression extracted from social annotation," in *Web Information Systems Engineering - WISE 2009*, 2009, pp. 401–414.

[4] H. Bhuiyan, J. Ara, R. Bardhan, and M. R. Islam, "Retrieving youtube video by sentiment analysis on user comment," in *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, 2017, pp. 474–478.

[5] M. Ueda, S. Yabe, D. Li, and S. Nakajima, "A cosmetic differences visualization system for beauty recommendation using the scores of various evaluation items," in *9th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems, IntRS 2022*, 2022, pp. 137–150.

[6] M. Ueda, Y. Taniguchi, Y. Matsunami, A. Okuda, and S. Nakajima, "Analyzing the tendency of review expressions for the automatic construction of cosmetic evaluation dictionaries," in *IAENG Transactions on Engineering Sciences*, 2020, pp. 81–95.

[7] T. Kim, H. Jo, Y. Yhee, and C. Koo, "Robots, artificial intelligence, and service automation (raisa) in hospitality: sentiment analysis of youtube streaming data," *Electronic Markets*, vol. 32, no. 1, pp. 259–275, 2022.

[8] K. Iizuka, T. Yoneda, and Y. Seki, "Greedy optimized multileaving for personalization," in *RecSys '19: Proceedings of the 13th ACM Conference on Recommender Systems*, 2019, pp. 413–417.

[9] M. P. Kato, T. Yamamoto, T. Manabe, A. Nishida, and S. Fujita, "Overview of the ntcir-13 openliveq task," in *Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies*, 2017.

[10] M. P. Kato, T. Manabe, S. Fujita, A. Nishida, and T. Yamamoto, "Challenges of multileaved comparison in practice: Lessons from ntcir-13 openliveq task," in *CIKM '18: Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 2018, pp. 1515–1518.

[11] K. Hofmann, S. Whiteson, and M. De Rijke, "A probabilistic method for inferring preferences from clicks," in *Proceedings of the 20th ACM international conference on Information and knowledge management*, 2011, pp. 249–258.