Topic Detection for Video Stream based on Geographical Relationships and its Interactive Viewing System

Itsuki Hashimoto
Kwansei Gakuin University
ewd16738@kwansei.ac.jp

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

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

Kazutoshi Sumiya
Kwansei Gakuin University
sumiya@kwansei.ac.jp

***Abstract*—During the recent years of Internet TV spread, researches on recommending relevant information for TV programs have been actively conducted. The NHK's Hybridcast provides a service that recommends relevant information on the same screen during the broadcast of a TV program. However,there is currently no service for recommending supplementary information based on the users' viewing behavior. Based on this research background, we first extract geographic words (location names) and topics of each scene using closed captions of TV programs. Next, we analyze the user's viewing behavior to extract the scenes selected by the user in the sequence. After that, we can detect the topics of the user's selected scenes. Therefore, the supplementary information is recommended by generating queries based on geographical relationships using geographical words and topics. In this paper, we discuss our proposed system for supporting interactive viewing of TV programs, which is based on the viewing behavior of users and geographic relationships.**

***Index Terms*—Video viewing support, TV programs, User interaction, Geographical relationships**

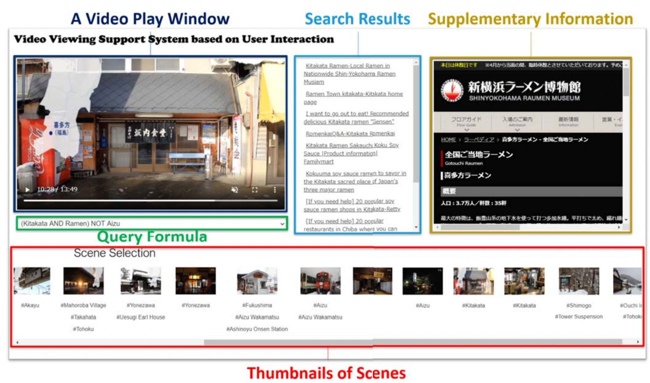

Fig. 1. Video viewing support system.

I. Introduction

Over the last years, with the spread of subscription services such as Netflix and Hulu, it has become easier to extract the user's tendency to view videos and to manage the video metadata. Therefore, researches are being actively conducted on the construction of program maps using the video metadata, and the recommendation of videos using the contents of videos [5], [6]. Moreover, research is being carried out to generate queries and implement video hyperlinks using video metadata, visual information, and textual information [7]. In addition, users can watch TV programs anywhere at any time. However, it leads fewer people to watch the whole program and more people to skip a few scenes in the TV program. Therefore, in this work, we propose a user interface that can allow users to select scenes as viewing behavior.

In previous works using TV programs, Wang et al. [1]–[3] proposed a method to extract location names in closed captions and recommend routes and images based on the semantic relationships between the location names. Kakimoto et al. [4] proposed a method to generate query keywords using geographical features of location names appearing in closed captions. These studies suggested that the use of geographical information is effective for recommending information in TV programs. That is, similar to our work, we use geographical relations between the location names that appear in each scene to determine what type of information to recommend. For example, there is a TV program that broadcasts "Tokyo" in the comparison scene (the previous viewing history) and "Tokyo University" in the corresponding scene (the watching scene). In this case, it can be expressed as "Tokyo $\ni$ Tokyo University". In this way, the information that is related to "Tokyo University in Tokyo" will be broadcasted. As mentioned above, it is possible to determine the information to recommend using geographical relationships between location names in scenes. The remainder of this paper is structured as follows. The next section provide an overview of our proposed video viewing support system and discuss previous research which has been carried out related to TV programs and geographical information. Section III explains our proposed topic extraction method. Section IV describes our proposed query generation method. Section V shows our evaluation experiments for the user interface and the query generation. Finally, Section VI concludes this paper and presents future works.

978-1-6654-1865-2/21/$31.00 ©2021 IEEE
DOI 10.1109/MIPR51284.2021.00012

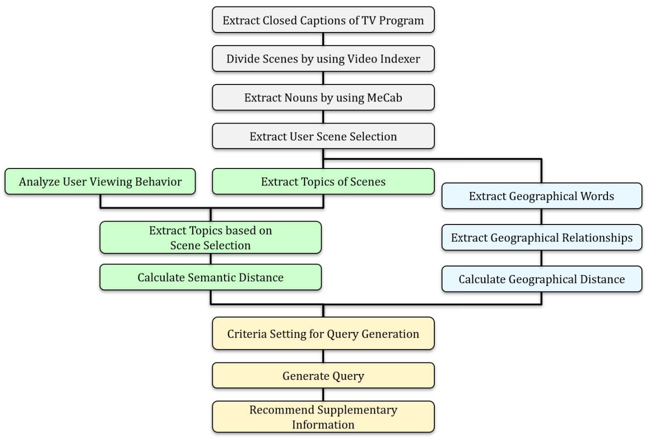

Fig. 2. The flow of our video viewing support system.

II. System Overview and Related Work

A. *Video Viewing Support System*

In this work, the goal is to recommend information according to the content of the TV program based on the user's viewing behavior. A user interface of our proposed video viewing support system is shown in Fig. 1, it can obtain the user's viewing behavior is used to extract the user's tendency and interests. We define the user's scene selection as the viewing behavior, and the topics and geographical words in the extracted scenes as the user's interests.

The user interface is divided into five parts. The video play window is shown in the upper left of the screen. The thumbnails of the scenes for the scene selection are shown at the bottom of the screen, and the location names appearing in each scene are presented as tags. The middle left of the screen shows the query expressions generated by the topics and the geographical relationships between scenes. The top 10 Google search results obtained by query formulation are shown in the upper center of the screen. On the upper right of the screen, the top web page of the search results is shown as supplementary information.

The flow of our video viewing support system is shown in Fig. 2. In the first pre-process, we extract closed captions of TV programs and divide scenes using Microsoft Video Indexer. Then, we extract nouns from the closed captions by the morphological analysis using MeCab. Next, we extract the scenes that were selected by the user through the user interface. As shown in the left side of Fig. 2, to extract topics, we extract words from the user's selected scenes based on the *LDA* method by considering the user's viewing behavior. After that, we calculate the semantic distance between the scenes and use it in the recommendation of the supplementary information. As shown in the right side of Fig. 2, to extract geographical relationships, we extract geographical keywords (location names) from each scene. After that, we calculate the geographical distance between the scenes using the geographical hierarchical structure and the real distance. Next, we generate the query formula based on the geographical relationships using both the semantic distance and the geographical distance. Finally, supplementary information is recommended on the user interface using these generated query formula.

B. *Related Work*

Recently, many research using TV programs have been performed. Additionally, as a research target of videos, there are few studies on the divination of scenes based on the changes of topics in the TV program. Baraldi et al. [8] proposed a method for dividing scenes by exploiting temporal changes in the topic existing in video frames. Sidiropoulos et al. [9] proposed a method for dividing scenes by using audio features, such as the BGM and the narration. Rasheed et al. [10] analyze the structure of movies and TV programs, using image features in addition to textual data, such as open captions and closed captions of video frames. In contrast, Liu et al. [11] focus on shots in a video, and use characters and captions that appear in the shots. Moreover, they also extracted voice information such as narrations and BGMs and proposed a method for dividing scenes by using these two data. In this work, we refer to the topic extraction and feature extraction as it is used for scene segmentation. Further, in the research of the use of video frames, research on video summarization has been actively studied. For video summarization, as in the case of scene division, features and entity changes are extracted from content in videos, and personalized short videos are recommended to users. Furthermore, in the research of using the user interaction of video, Salim et al. [12] implemented a system that makes it easy to know the contents of video. For that purpose, five types of supplementary information were generated using the modal features of the video, and tabs were placed on the same screen on the video. This allows users could select which supplementary information they wanted to view. This study confirmed the effectiveness of the action of displaying supplementary information on the same screen and operating it by users. Also, Olsen Jr. et al. [13] analyzed the user DOI (Degree of Interest Function) using the American football game video and summarized the video. This study showed that it is effective to use user operations such as skipping and rewinding in video to extract user interests and viewing habits. Since our work utilizes the user's scene selection, it is definitely close in that he/she does not see the entire video. But the video summarization is definitely difficult because it can break the commitment and intention of the creator of the video [14]–[16]. Additionally, research on recommending information related to shopping and museums has been also conducted with the help of videos. As such, research focused on videos of a certain category helps determine future policies [17], [18]. Since our work includes content related to geographical information, we also refer to research other than video-based research. Kitayama et al. [19], Kobayashi et al. [20], and Kurata [21] proposed methods for recommending geographical information. In these studies, feature vector extraction is essential for recommending geographical information. Kurata [21] uses Topological Rela-

TABLE I
TOPIC EXTRACTION USING *LDA-U*, *LDA*, AND *TFiDF*

TV Program	Scene Selection	LDA-U	LDA	TFiDF
A	7 → 9 → 13	Becky, back, sudden, first, Kagetsutei,seafood, hot spring	Becky, back, sudden, first, Kagetsutei, seafood, book	certificate, incinerator, final destination, reservation
B	12 → 16 → 22	Arima, hot spring, town, seawater, research, together, fluffy	Arima, hot spring, town, scent, seawater, research, center, together, fluffy, hot water, earlier	expressway, main street
C	4 → 3 → 5	birds, Fujimae, food, mud, tidal flat, nutrients, location, snipe, feathers	birds, Fujimae, food, mud, tidal flat, nutrients, location	fish

TABLE II
TV PROGRAMS USED IN THE PRELIMINARY EXPERIMENT

TV program	Name	Time	#Scenes
A	Soda tabi ni iko	46 min	22
B	Buratamori #97	45 min	28
C	Sawayaka shizen hyakei #2	15 min	8

tionships in the recommendation of geographical information. In our study, these relations are considered to be effective, and they are used to extract geographical relationships.

III. TOPIC EXTRACTION

In this work, query formulae are generated using topics and geographical keywords when recommending supplementary information. For this purpose, we explain the extraction method of topics used in query formulae in the relevant scenes.

A. Extracting Topics of Scenes

Latent Dirichlet Allocation (*LDA*) [23] is used as the topic extraction method, and the keywords appearing in the program are weighted. This makes it possible to determine which keywords are topics in each scene. The following is a description of the specific process flow. We define a shot as a camera change in a program, a scene as a semantic set of shots, and a narration or a line said in each shot as a document. The set of shots in the whole program is used as input for morphological analysis by MeCab, and only the nouns obtained from the analysis are used as keywords. In order to remove unnecessary words in the morphological analysis, we set one number or one hiragana character as Stopwords. Then, the frequency of occurrence of the extracted keywords is used for vectorization using *TFiDF*, and the topic distribution is obtained using *LDA*.

The hyperparameters α and β were set to α = 50/T and β = 0.01 according to Zhao et al. [24] The number of topics was set according to Watanabe et al. [25]. The number of topics was run from 5 to 30 in increments of 5, and the number of topics judged to be the most appropriate after manual inspection was used. Since the appropriate number of topics depends on the length of the program, the number of topics = 15 for a 45-minute program and the number of topics = 10 for a 15-minute program.

To determine the topics for each scene, the top three topics for each scene are extracted using the average of the probabilities of belonging to each topic for each shot in the scene. Then, the top 10 keywords in the top 3 topics are extracted, and the average value of the probability that each keyword belongs to the topic is set as the threshold value, and the keywords above the threshold value are set as topic words.

B. Extracting Topics based on User Viewing Behavior

Since the purpose of this study is to recommend supplementary information considering the user's viewing operation, we propose *LDA-U* as a method that also considers the viewing operation in topic extraction. In *LDA-U*, as in *LDA*, a set of shots from the entire program is used as input. However, when determining the topic for each scene, the average value of the probability of belonging to each topic for each shot is calculated using the corresponding scene and two scenes from the previous scene selection history.

In addition, when calculating the average value, a constant is assigned as a weight. We multiply the probability of the corresponding scene by 0.5, the probability of the previous scene by 0.3, and the probability of the previous scene by 0.2. This makes it possible to construct a topic distribution that takes into account the topics in the viewing history.

C. Comparison of topic extraction results for each method

The results of the comparison of topic words extracted by *LDA-U* and *LDA* using actual programs are shown in Table II. The results of the conventional method *TFiDF* are also shown as a comparison method. The calculation of *TFiDF* is based on the following equation, where the total number of scenes in the entire program is N, the corresponding scene is defined as s, and the keywords appearing in the corresponding scene are defined as k.

$$\mathit{TFiDF} = TF(k,s) \times log\frac{N}{DF(k)} \tag{1}$$

The results of the topic extraction using the three methods are shown in Table I. As a result, topic extraction using *LDA-U* for TV programs A and C can extract topic words such as ”hot spring” and ”snipe” that were extracted as topic words in the previous scene. In addition, when comparing the topic words extracted by *TFiDF* with those extracted by *LDA-U* and *LDA*, *TFiDF* is calculated using only the frequency of occurrence, so that only the keywords with high frequency of occurrence in the corresponding scene are extracted, while *LDA-U* and *LDA* can extract relevant keywords that do not appear in the corresponding scene. From the above results, it is confirmed that the proposed method *LDA-U* is capable

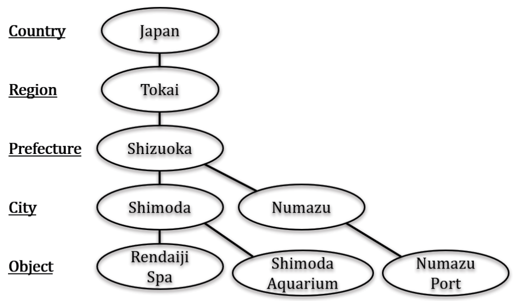

Fig. 3. Example of geographical hierarchy.

of extracting topics in consideration of the user's viewing operation.

IV. Query Generation

In this work, we use web pages to recommend supplementary information. Therefore, we generate a query to search the web pages. To generate the query, we use the Boolean search model, which is also used in information retrieval.

A. Criteria Setting for Query Generation

In the generation of queries using geographical relationships and topics between scenes, the contents between scenes can be definitely different, or the geographical keywords can be definitely far apart from each other. Therefore, it is necessary to determine the conditions before generating a query. In this work, we define two distances: the geographical distance that uses the geographical hierarchical structure and the real distance between geographical keywords based on the geographical relationships between scenes, and the semantic distance that uses the similarity of the contents between the scenes. The criteria to generate the query formula based on the geographical relationships using both the semantic distance and the geographical distance is shown in Table III.

1) Calculating Geographical Distance: The geographical distance is calculated using the geographical hierarchical structure and the real distance. Fig. 3 shows an example of the geographical hierarchical structure. The calculations are performed by taking into account the range of geographical keywords and their semantic relationships, using a five-level hierarchy. Moreover, in the case of a wide range of geographical keywords, the real distance is calculated using the center point. The formula used to calculate the geographical distance *GeoD* is as follows:

$$\begin{aligned} GeoD(k_i, k_j) &= GeoH(k_i, k_j) * D(k_i, k_j) \\ GeoH(k_i, k_j) &= |pos(k_i) - pos(k_j)| \\ D(k_i, k_j) &= |lat(k_i) - lat(k_j)| + |lon(k_i) - lon(k_j)| \end{aligned} \quad (2)$$

Here, $GeoD(k_i,k_j)$ returns the geographical distance between the geographical keywords k_i and k_j. $GeoH(k_i,k_j)$ returns the weight of the semantic relationships between k_i and k_j in the geographical hierarchy, *pos* denotes the position. In the geographical hierarchy, the values are assigned to the position as follows: Country:5>Region:4>Prefecture:3>City:2>Object:1. In this way, these values are used to give weight to semantic relationships between the geographic keywords. $D(k_i,k_j)$ returns the real distance between k_i and k_j, *lat* denotes the latitude and *lon* denotes the longitude.

2) Calculating Semantic Distance: To calculate the semantic distance, we compute the content similarity between the scenes. It will determine the movement of related scenes and avoid the recommendation of information between non-related scenes. To do so, we compute the Euclidean distance using keywords of the corresponding scene and compare the scenes. The *TFiDF* value for each keyword is used for the feature vector. The formula used to calculate the semantic distance *SemD* is as follows:

$$\begin{aligned} SemD(s_i, s_j) &= \sqrt{\sum_{i=1}^{N}(s_j - s_i)^2} \\ s_i &= \{TFiDF(k_1), TFiDF(k_2), ..., TFiDF(k_j)\} \end{aligned} \quad (3)$$

Here, $SemD(k_i,k_j)$ returns the semantic distance between the scenes s_i and s_j.

B. Query Generation based on Geographical Relationships

In this work, we use the geographical relationships between scenes based on the viewing order obtained from the user's viewing history. We refer to the four types of Topological Relationships [22] as geographical relationships: **Disjoint**, **Equal**, **Inside**, and **Contain**. Then, we determine the types of supplementary information according to the geographical relationships as follows:

Disjoint: Comparing the information on regions is recommended because the regions are far away from each other.

Equal: Comparing the information on the topics is recommended because the topics are in the same regions

Inside: The detailed information of a region is recommended since the region is also included in the other region.

Contain: The summary information of a region is recommended since the region contains the other region.

Fig. 4 shows an example of each geographical relationship, and Table IV shows an example of queries that are generated by each geographical relationship. When generating a query formula, multiple geographical keywords may appear for one scene, therefore it is necessary to determine the ranking of the query formulae. The query formulae generated using a pair of geographical keywords with the closest geographical distance are shown at the top. Additionally, users can also select the query formulae through the user interface.

V. Evaluation

In this section, we evaluate the usability of the user interface and the effectiveness of our proposed query generation by the subject evaluation. In the user interface evaluation experiment, the usability of the user interface is evaluated using SUS

TABLE III
THE CRITERIA FOR QUERY GENERATION BASED ON GEOGRAPHICAL RELATIONSHIPS USING GEOGRAPHICAL AND SEMANTIC DISTANCES

Geographical Relationship	Disjoint	Equal	Inside	Contain	Disjoint	Equal	Inside	Contain
Geographical Distance / Semantic Distance	Near				Far			
Near	Generate	Generate	Generate	Generate	Generate	Not Generate	Generate	Generate
Far	Generate	Generate	Generate	Generate	Not Generate	Not Generate	Not Generate	Not Generate

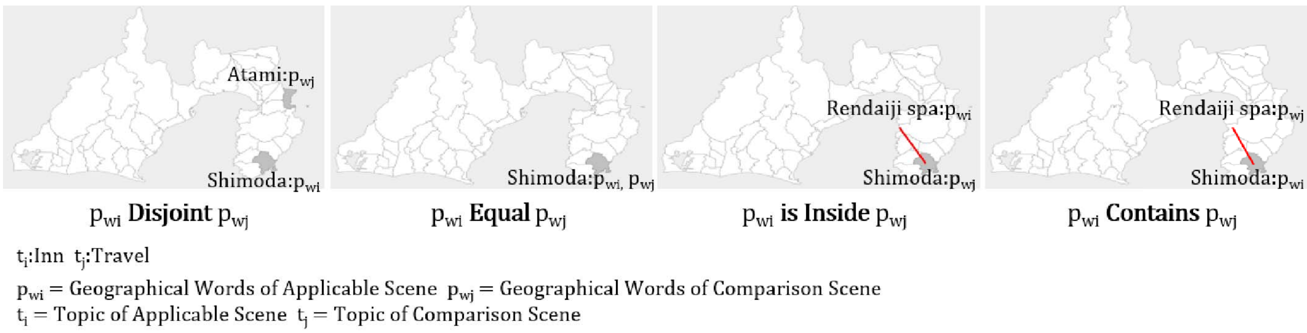

Fig. 4. Example of geographical relationships.

TABLE IV
QUERY GENERATION BASED ON GEOGRAPHIC RELATIONSHIPS

Geographic Relationship	Recommendation	Query	Example of Query
(1) Disjoint	Comparison of Regions	$(p_{w_i}\ AND\ t_i)\ OR\ (p_{w_j}\ AND\ t_i)$	(*Shimoda AND Inn*) *OR* (*Atami AND Inn*)
(2) Equal	Comparison of Topics	$(p_{w_i}\ AND\ t_i)\ OR\ (p_{w_i}\ AND\ t_j)$	(*Shimoda AND Inn*) *OR* (*Shimoda AND Travel*)
(3) Inside	Details of Regions	$(p_{w_i}\ AND\ t_i)\ NOT\ p_{w_j}$	(*Rendaiji Spa AND Inn*) *NOT Shimoda*
(4) Contain	Summary of Regions	$(p_{w_i}\ AND\ t_i)\ NOT\ p_{w_j}$	(*Shimoda AND Inn*) *NOT Rendaiji Spa*

Score. In the evaluation experiment of query generation, the evaluation by the student subjects is used to determine whether the proposed method can extract effective supplementary information.

A. Experiment 1: Usability of User Interface

The user interface of our prototype system is shown in Fig. 1. Fig. 5 shows an example of operations on the user interface. For example, the user can select scenes on the left screen, the supplementary information is presented based on the geographical relationships between the scenes, as shown on the right screen. After clicking on the video, the user can play the video on the top-left screen. Furthermore, in this experiment, the same user interface is used without scene selection. The question items of the SUS score are as follows:

- Q1 I think that I would like to use this user interface frequently.
- Q2 I found the user interface unnecessarily complex.
- Q3 I thought the user interface was easy to use.
- Q4 I think that I would need the support of a technical person to be able to use this user interface.
- Q5 I found the various functions in this user interface were well integrated.
- Q6 I thought there was too much inconsistency in this user interface.
- Q7 I would imagine that most people would learn to use this user interface very quickly.
- Q8 I found the user interface very cumbersome to use.
- Q9 I felt very confident using the user interface.
- Q10 I needed to learn a lot of things before I could get going with this user interface.

The subjects are asked questions using a 5-point Likert scale. The adjective ratings for the SUS Score are the following: A) Excellent: >80.3, B) Good: >68.0-80.3, C) Okay: >68.0, D) Poor: >51.0-68.0, E) Awful: >51.0. In this experiment, we perform two user interface patterns using the SUS score, with and without scene selection, and each pattern is evaluated by nine subjects and 18 subjects in total. The average values of the SUS Score for each question and each pattern are shown in Table V.

From the results of Table V, we found that the grades of the average SUS scores for the user interface with scene selection is *Okay*, and without is *Poor*. Looking at the average rating score for each question in Table V, we could confirm that the user interface with scene selection (proposed) is easy to use. In Q3, Q7, and Q9, many subjects answered that they could use it easily and that they could easily understand how to use it. In Q5 and Q6, many subjects answered that the content and navigation of the user interface were meaningful, in the sense of unity. In Q1, many subjects answered that they would like

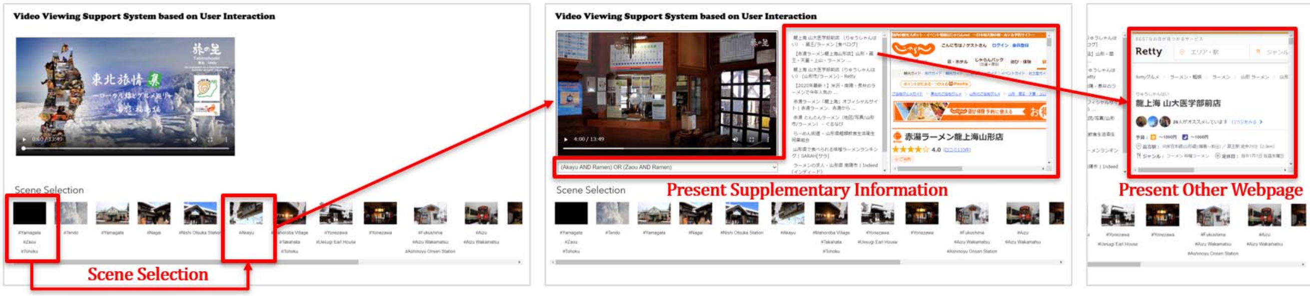

Fig. 5. Example of operations on the user interface.

TABLE V
AVERAGE OF SUS SCORE FOR USER INTERFACE

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	SUS Score	Grade	Adjective Rating
With scene selection	3.7	2.7	3.9	2.8	4.2	1.9	4.1	2.0	4.0	2.9	**69.2**	C	Okay
Without scene selection	3.2	2.9	3.7	2.0	4.1	2.3	3.3	2.0	3.8	3.1	64.4	D	Poor

TABLE VI
TV PROGRAMS FOR QUERY GENERATION EVALUATION EXPERIMENT

TV Program	Title	Time Length	Scenes	Percentage	Respondents	Scene Selection
Travel Program	Tabi no hoshi	14m	17	24%	11	$2 \rightarrow 6 \rightarrow 11 \rightarrow 13 \rightarrow \mathbf{14}$
News Program	News kyo-ichiniti	15m	29	7%	11	$1 \rightarrow 2 \rightarrow 6 \rightarrow 7 \rightarrow 13 \rightarrow \mathbf{19}$

TABLE VII
RESULTS OF WEB PAGES OBTAINED BY THREE METHODS

Method	User Rating (Travel)	User Rating (News)
Proposed Method	**3.56**	**3.00**
TFiDF	2.38	2.75
TF	1.76	2.36

to use it frequently, and the average rating score is 3.7. For that reason it was confirmed that there are many users who want to frequently use the user interface used in this research.

B. Experiment 2: Effectiveness of Query Generation

In this experiment, we verified the effectiveness of our proposed method for query generation by comparing it with the methods *TFiDF* and *TF*. For this, we obtained the top 10 web pages by each method for two TV programs: a travel program and a news program in Table VI.

In the travel program, the query formula **(***Kitakata AND Ramen***)** *NOT Aizu* generated from the corresponding scene (Scene 14) and the comparison scene (Scene 13) by our proposed method, the query keyword "Kitakata" obtained by *TFiDF*, and the query keyword "Ramen" obtained by *TF*. In the news program, the query formula **(***China AND Escort warship***)** *OR* **(***USA AND Escort warship***)** generated from the corresponding scene (Scene 19) and the comparison scene (Scene 13) by our proposed method, the query keyword "Escort warship" obtained by *TFiDF*, and the query keyword "Ministry of defense" obtained by *TF*.

Table VII shows the average user ratings for the travel and news programs by each method using a 5-point Likert scale. From the results of Table VII, we confirmed that the web pages as supplementary information obtained by our proposed method are more effective than the methods *TFiDF* and *TF*. Moreover, we could confirm that our proposed method is effective for TV programs in different genres.

VI. CONCLUSION

In this paper, we have proposed a video viewing support system based on user viewing behavior. We generated query formulae using geographical relationships and topic words to recommend supplementary information. For this, we used the geographical hierarchical structure to calculate the geographical distance. Considering that the processing speed will be slow, we plan to use the zip code for computation in the future. Finally, we evaluated the usability of the user interface and the effectiveness of the generated query formulae, and we obtained higher results than the compared methods.

In the future, this method uses geographical keywords as feature words. However, we consider that it can also be used in E-learning by changing the feature words. Therefore, we are considering of conducting an experiment using videos of another genre. In addition, we consider that it can be used for local programs that focus on a narrow area. Therefore, we are considering of conducting experiments with such videos.

ACKNOWLEDGMENT

The research was partially supported by ROIS NII Open Collaborative Research 2020-(20FC04), and JSPS KAKENHI Grant Number JP19H04118.

REFERENCES

[1] Yuanyuan Wang, Daisuke Kitayama, Yukiko Kawai, and Kazutoshi Sumiya, "Automatic Street View System Synchronized with TV Program using Geographical Metadata from Closed Captions", in Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces (AVI 2014), pp. 383-384,May 27-30, 2014.

[2] Yuanyuan Wang, Daisuke Kitayama, Yukiko Kawai, Kazutoshi Sumiya, and Yoshiharu Ishikawa, "An Automatic Video Reinforcing System for TV Programs using Semantic Metadata from Closed Captions", in International Journal of Multimedia Data Engineering and Management, Volume 7, Issue 1, pp. 1-21, January-March 2016.

[3] Yuanyuan Wang, Maho Nishizawa, Yukiko Kawai, and Kazutoshi Sumiya, "Location-based Image Viewing System Synchronized with Video Clips", in Proceedings of the 13th International Conference on Location-Based Services, LBS 2016, pp. 233-238, November 14-16, 2016.

[4] Honoka Kakimoto, Toshinori Hayashi, Yuanyuan Wang, Yukiko Kawai, and Kazutoshi Sumiya, "Query Keyword Extraction from Video Caption Data based on Spatio-Temporal Features", in Proceedings of the International MultiConference of Engineers and Computer Scientists 2018, Volume I, IMECS 2018, pp. 405-408, March 14-16, 2018.

[5] Xingzhong Du, Hongzhi Yin, Ling Chen, Yang Wang, Yi Wang, and Xianfang Zhou, "Personalized Video Recommendation Using Rich Contents from Videos",in Proceedings of the 2020 IEEE Transactions on Knowledge and Data Engineering, Volume:32, Issue: 3, March 1, 2020.

[6] Junyu Gao, Tianzhu Zhang, and Changsheng Xu, "A Unified Personalized Video Recommendation via Dynamic Recurrent Neural Networks",in Proceedings of the 2017 ACM Multimedia ,ACMMM'17, October 23-27, 2017.

[7] Mohammad Reza Kavoosifar, Daniele Apiletti, Elena Baralis, Paolo Garza, and Benoit Huet, "Effective video hyperlinking by means of enriched feature sets and monomodal query combinations", International Journal of Multimedia Information Retrieval volume 9, pages215–227(2020).

[8] Lorenzo Baraldi, Costantino Grana, and Rita Cucchiara, "Scene Segmentation using Temporal Clustering for Accessing and Re-using Broadcast Video", in Proceedings of the 2015 IEEE International Conference on Multimedia and Expo (ICME 2015), 6 pages, June 2015.

[9] Panagiotis Sidiropoulos, Vasileios Mezaris, Ioannis Kompatsiaris, Hugo Meinedo, Miguel Bugalho, and Isabel Trancoso, "Temporal Video Segmentation to Scenes Using High-Level Audiovisual Features", in IEEE Transactions on Circuits and Systems for Video Technology, Volume 21, Issue 8, pp. 1163–1177, 2011.

[10] Z. Rasheed and M. Shah, "Detection and Representation of Scenes in Videos", in IEEE Transactions on Multimedia, Volume 7, Issue 6, pp. 1097–1105, December 2005.

[11] Cailiang Liu, Dong Wang, Jun Zhu, and Bo Zhang, "Learning a Contextual Multi-Thread Model for Movie/TV Scene Segmentation", in IEEE Transactions on Multimedia, Volume 15, Issue 4, pp. 884-897, June 2013.

[12] Fahim A. Salim, Fasih Haider, Owen Conlan1, and Saturnino Luz, "An Approach for Exploring a Video Via Multimodal Feature Extraction and User Interactions", Journal on Multimodal User Interfaces (2018) 12:285–296.

[13] Dan R. Olsen Jr., and Brandon Moon, "Video Summarization Based on User Interaction", EuroITV'11, June 29–July 1, 2011, Lisbon, Portugal.

[14] Shayok Hakraborty, Omesh Tickoo, and Ravi Iyer, "Adaptive Keyframe Selection for Video Summarization", in Proceedings of the 2015 IEEE Winter Conference on Applications of Computer Vision (WACV 2015), pp. 702–709, 2015.

[15] Ba Tu Truong, and Svetha Venkatesh, "Video abstraction: A systematic review and classification", in Proceeding of ACM Transactions on Multimedia Computing, Communications, and Applications, February 2007.

[16] Aidean Sharghi, Boqing Gong, and Mubarak Shah, "Query-Focused Extractive Video Summarization", European Conference on Computer Vision (ECCV, 2016) pp 3-19

[17] Fausto C. Fleites, Haohong Wang, and Shu-Ching Chen, "Enabling Enriched TV Shopping Experience via Computational and Temporal Aware View-Centric Multimedia Abstraction", in IEEE Transactions on Multimedia, Volume 17, Issue 7, pp. 1068-1080, July 2015.

[18] Ben Rydal Shapiro, Rogers P. Hall, and David A. Owens, "Developing and using interaction geography in a museum", International Journal of Computer-Supported Collaborative Learning volume 12, pages377–399(2017).

[19] Daisuke Kitayama, Setsuko Miyamoto, and Kazutoshi Sumiya, "A Customizing Method of Digital Maps Based on User's Operations and Object's Appearing Patterns [in Japanese]", in Transactions of Information Processing Society of Japan (TOD), Volume 3, Number 4, pp. 65-81, October 2010.

[20] Kaori Kobayashi, Daisuke Kitayama, and Kazutoshi Sumiya, "Cinematic Street: Automatic Street View Walk-through System Using Characteristics of Modified Maps", in Proceedings of the 10th International Symposium on Web & Geographical Information Systems (W2GIS 2011), pp. 142–158, 2011.

[21] Yohei Kurata, "An overview of the research on topological relations and future issues in GIScience [in Japanese]", in Theory and Applications of GIS, Volume 18, Number 2, pp. 41–51, 2010.

[22] Max J Egenhofer, "A model for detailed binary topological relationships", Geomatica, 47, 3-4, pp.261-273, 1993.

[23] David M. Blei, Andrew Y. Ng, and Michael I. Jordan, "Latent Dirichlet Allocation", Journal of Machine Learning Research 3, pp. 993-1022, 2003.

[24] Wayne Xin Zhao, Jing Jiang, Jianshu Weng, Jing He, and Ee Peng Lim, "Comparing Twitter and Traditional Media using Topic Models", Proc. 33rd European Conference on Advances in Information Retrieval, pp. 338-349, 2011.

[25] Keita Watanabe, and Shohei Kato, "Tweet Recommendation System Reflecting User Preference Based on Latent Dirichlet Allocation and Collaborative Filtering", Proc. The 28th Annual Conference of the Japanese Society for Artificial Intelligence, 2014.