

Query Keyword Generated by Spatio-Temporal Features for Datacasting Services

Honoka KAKIMOTO[†], Toshinori HAYASHI[†], Yuanyuan WANG^{††}, Yukiko KAWAI^{†††}, and

Kazutoshi SUMIYA[†]

[†] Kwansei Gakuin University 2-1 Gakuen, Sanda-shi, Hyogo, 669-1337 Japan

^{††} Yamaguchi University 2-16-1 Tokiwadai, Ube-shi, Yamaguchi, 775-8611 Japan

^{†††} Kyoto Sangyo University Kamigamo-motoyama, Kita-ku, Kyoto, 603-8555 Japan

E-mail: [†]{des44337,den82687,sumiya}@kwansei.ac.jp, ^{††}y.wang@yamaguchi-u.ac.jp,

^{†††}kawai@cc.kyoto-su.ac.jp

Abstract Recently, people have begun searching for relevant information of each scene of TV program videos with other devices such as smartphones and tablets. While they view TV programs, users' interests change by each scene of the video. When they try to get information related to the content of the scenes, users have to input appropriate query keywords for a Web search. However, it takes users time and effort to find their requested information. Although some data casting services suggest related information to TV programs, the related information does not synchronize enough with each scene of the videos. To solve this problem, our system proposes a novel query keyword extraction method for Web searches, based on spatio-temporal features of videos using location names in the video caption data. We first extract all location names from closed caption for detecting the main topic of the scene. Then classify them into two distances: geographical distance between locations and the the relation of location names based on Wikipedia tree structure. Next, it determines main topics from all nouns to generate web search queries based on the appearance frequency. Therefore, suitable web pages for each scene can be found based on the generated query keywords through our system.

Key words closed caption, geographical relationships, recommender system, topic extraction.

1. Introduction

Recently, people have begun searching for relevant information for each scene on TV programs with other devices such as smartphones and tablets. While view TV programs, users' interests change with each scene of the video. When users want to get information related to the contents of the scenes, they have to input appropriate query keywords a for Web search. However, it takes users time and effort to find their requested web pages until they achieve the relevant information. Moreover, there are certain users, including children and elderly people, who are not able to input appropriate query keywords. Further, it is difficult to search various types of information through the Web at once.

Some datacasting services such as NHK Hybridcast [1] and other viewer participating program services recommend related information to TV programs on the interface. However, the recommended information does not synchronize with each scene of the videos. Therefore, it is necessary to recommend information related to each scene and users'

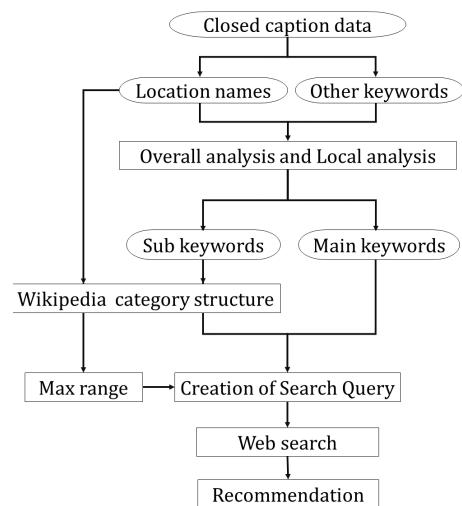


Figure 1 System Flow

concerns. TV programs are often associated with closed captions, and many researchers proposed systems utilizing topics in closed caption data of videos. In this work, we develop an automatic location-based recommendation sys-

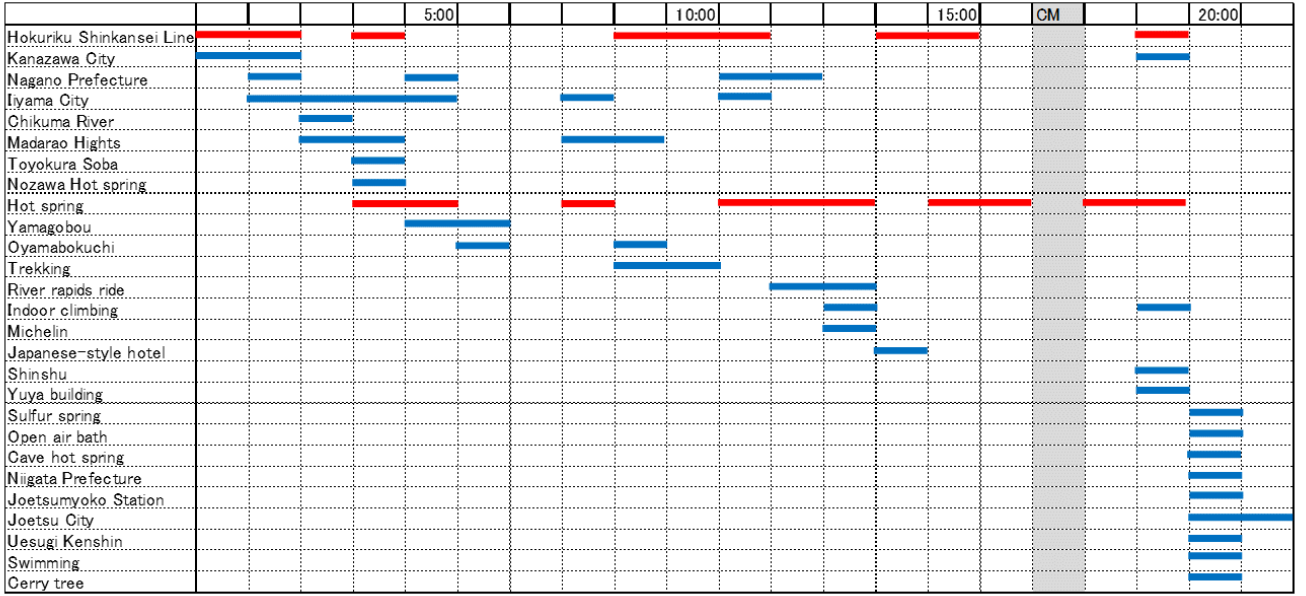


Figure 2 Location names and keywords related to tourism in closed caption data

tem using the concept of the automatic location-based image viewing system synchronized with video clips [2] by adding a real geographical distance and length of video times to generate web search queries.

2. System Overview and Related Work

2.1 System Overview

Figure.1 shows the system flow of our proposed method for generating web search query based on location names in the closed caption.

First, the system extracts location names in closed captions from an MPEG file of a TV program based on the method of related work. It classifies location names and deletes unnecessary words as an outlier, and sets the maximum range of area that the system recommends based on the location names in closed caption data. Second, the system selects the main keywords and sub keywords on the TV program in two analyses, based on the time length. Then it creates a web search query by combining the main keywords and sub keywords. Third, the system searches suitable web pages for the scenes with the search query. Then, it recommends several web pages on the user interface, which are detailed information and related information by each scene. Web pages are recommended as three tabs of user interface: “go”, “eat,” and “buy.” An example of system flow is described below:

A TV program related to the tourist spot along Hokuriku Shinkansen Line [6] has the following flow:

- (1) Extract location name: “Niigata,” “Kanazawa,” etc.
- (2) Max range: Hokuriku area

- (3) Select “Hokuriku Shinkansen Line” as the main keyword.
- (4) Select “Hot spring,” “Skiing,” etc. as sub keywords.
- (5) Create search query: (Skiing or Hiking) and Hokuriku Shinkansen Line.
- (6) Search web pages with the queries.
- (7) Recommend web pages about skiing ground or hot spring along Hokuriku Shinkansen Line on the tab “go” of the user interface.

2.2 Related Work

Nishizawa et al [2] extracted a semantic structure of location names in closed caption data by utilizing Wikipedia categories and detected relevant topics of location names in the semantic structure. In our work, we extract a semantic structure of location names in closed caption data based on their method. In addition, they proposed location-based image viewing system synchronized with video clips. In their work, the system recommends the images and maps information related to the scenes based on location names in closed caption data of travel video clip. To recommend more suitable information for users’ interests in each scene, we recommend contents from the web pages of e-commerce, travel, and restaurant search sites.

Wang et al [3] proposed an automatic video reinforcing system based on popularity rating of scenes and level of detail controlling scenes based on closed caption data. They proposed a novel automatic video reinforcing system with a media synchronization mechanism and a video reconstruction mechanism based on closed caption data. Their proposed system recommends some web contents such as YouTube video clips and images related to the scene of a video clip.

To recommend suitable information for a travel TV program, our proposed system recommends web pages related to tourism and local specialties.

Son [4] proposed the system that segment a broadcasting content into semantic units, scenes, based on its multiple characteristics. In this work, they analyzed scenes and generates their keywords, topics, and stories.

Ma et al [5] proposed the system web pages related to the TV-program content are retrieved automatically in real time. In our work, we propose the automatic recommendation system like their system.

3. Query Keyword Extraction based on Spatio-Temporal Features

3.1 Extraction and Classification of Location Names

Our proposed system extracts location names and other keywords from all nouns in the closed captions from an MPEG file of TV program based on the method of related work [2]. Then, it classifies location names into two factors: geographical distance between locations and semantic distance between location names.

First, it deletes the location names that extremely deviate from a group of other location names in a closed caption. Subsequently, we sets a maximum range of recommendation with the rest of the names. Second, to recommend relevant web pages effectively, the system sets a maximum range of the recommendation of web pages. We measure the semantic distances from location to location by using the semantic structure based on the Wikipedia category structure for the creation of web a search query, as described in detail in section 3.3.

3.2 Determination of Main Topics and Sub Topics

With the keywords in closed captions, this system selects main topics, sub location names and sub topics on the TV program to generate the web search query. Figure.2 shows the appearance of location names and keywords related to tourism extracted from closed caption data of a 20-min TV program [6].

Here, closed captions of TV commercials are not extracted. The first row of Figure.2 shows the time sequence. Red lines show keywords that appear periodically in the TV program, and blue lines show other keywords. The keywords that appear periodically in closed captions are determined as main location names and sub location names, and other keywords are determined as sub topics based on the length of time.

In this work, we make a ranking of location names in the overall analysis. The main objective of the overall analysis is to defines the location name that appear most periodically in the closed captions as the main location name based

on frequency and average of interval. And it is the word that present the main theme or the atmosphere of the TV program potentially. Based on this analysis, it selects more than one location name for recommendation related to the location in the closed caption. In addition, other location names in high ranks are defined as sub location names. We weight the proper nouns, location names, landmarks, and foods based on the result of the preliminary experiment: section 4. Therefore, our proposed system increases the priority of main topics to enhance the accuracy of Web search. It means that semantic interpretation of the main topics changes based on each scene of the video.

Thus, the system creates a various query for Web search. In addition, it deals with keywords that appear for short time as the local analysis. To generate search query keyword to recommend detail information, it combines the overlapping topics. Similar to overall analysis, main topics also take priority in the local analysis.

3.3 Generation of Web Search Query

To search web pages related to the scenes, this system creates a query in two ways based on the location names and other keywords that are classified: AND search and AND-OR search. First, it generates AND search query with the main location name and some sub topics based on local analysis to search and recommend detail information. Second, it generates AND-OR search query with the main location name and other location names and some sub topics based on the overall analysis. Other location names and sub topics are in parallel relationships based on Wikipedia category structure. Finally, the system searches web pages related to each scene of TV program from travel, e-commerce, and restaurant search sites based on the search queries. Then, the results of the search are recommended.

3.4 Recommendation of Web Pages

Web pages searched with a search query are recommended on three tabs of the user interface as shown in Figure. 3. At least three web pages are recommended on each tab. Tab “go” displays the web pages related to tourism and events. Tab “buy” and “eat” display the web pages that are searched from e-commerce and restaurant search sites. Users can change the tabs based on their interests. To enhance operability, this work assumes the use of the system on a tablet. Users tap on a thumbnail image and browse the web pages on a browsing window while watching the TV program on a video player window.

4. Preliminary Experiment

We conducted a questionnaire survey with several short video clips of TV program related to the tourism [7] [8] [9], and the group of all nouns in the closed caption data. The

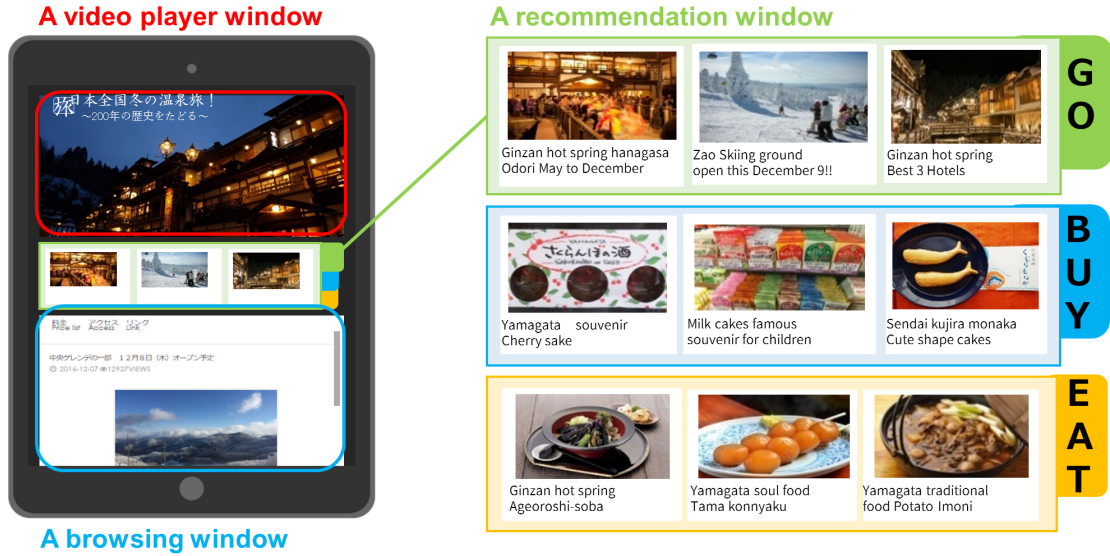


Figure 3 User Interface

purpose of the survey is to analyze the trend of combinations of the keywords that viewers use while watching TV programs. Each video is divided based on the scene changes. To analyze the trend of search queries for relevant information to the scenes, viewers add at least one keyword to make a search query. After respondents watch several videos, they answer the following questions:

Q1: Please circle all keywords you are interested in.

Q2: Please make some web search keywords with the keywords that you circled in Q1 and at least 1 keyword that you want to add.

The result of Q1 shows 40% of all circled nouns were location names, landmarks, and foods. In addition, more than a quarter of them are the proper nouns. It shows that viewers are expected to be interested in these type of nouns. Therefore, we weight the proper nouns, location names, landmarks, and foods in the determination of main topics and sub topics: section 3.2. In addition, to research on the effects of advance knowledge, we asked respondents whether they know the locations and landmarks in the scenes in advance or not. The result shows that viewers who have advance knowledge tend to search relevant information than detailed information, and the respondents who do not have advance knowledge tend to search detailed information. To recommend suitable web pages for viewer individual's property, we reflect the results in the recommendation method.

The result of Q2 shows only 17 % of web search keywords for relevant search contain location names and landmarks. It shows implicit needs of respondents, and to recommend relevant information searched with location names and landmarks do not appear in the closed caption is effective, because this way of making web search query is an idea that

doesn't exist in viewers thinking. Therefore, we consider to recommend relevant information with location names and landmarks that do not appear in closed caption can expand viewers interests in the scenes. In addition, over 50% of web generated web search queries are for detailed search, it shows viewers explicit needs.

5. Conclusion

In this paper, we proposed a query keyword extraction method for Web search based on location names in the closed caption data of videos. First, our proposed system extracts the location names and classified them into two factors: geographical distance between locations and semantic distance between location names. Then, it determined main topics and sub topics based on length of time. Next, it generates a web search query by combining main topics and sub topics, and the suitable web pages for scenes were found based on the generated web search query. Thus, it recommends detail information and related information through the user interface of our proposed system.

As future work, we are planing to evaluate our system with a questionnaire survey. For this, we plan to make several demonstration videos with TV programs about travel and tourism. In addition, we plan to construct a dictionary to extract appropriate keywords to TV program related to travel and tourism. Another future direction is to recommend information with various types of media such as SNS, videos, reviews, etc. to expand the users' interests.

References

- [1] NHK Hybridcast (2013) Accessed 7 September 2017. Retrieved from <http://www.nhk.or.jp/hybridcast/online/>
- [2] Y. Wang, M. Nishizawa, Y. Kawai, K. Sumiya, "Location-

based Image Viewing System Synchronized with Video Clips ”, Proc. of the 13th International Conference on Location Based Services (LBS 2016), p. 233-238, Vienna, Austria, November, 2016.

- [3] Y. Wang, Y. Kawai, K. Sumiya, Y. Ishikawa, “ An Automatic Video Reinforcing System based on Popularity Rating of Scenes and Level of Detail Controlling”. Proc. of the 2015 IEEE International Symposium on Multimedia (ISM 2015), pp. 529-534, 2015.
- [4] Jeong-Woo Son, Wonjoo Park, Sang-Yun Lee, Jinwoo Kim, Sun-Joong Kim, ”Smart Media Generation System for Broadcasting Contents”. Proc. of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2017), pp. 1297-1300, 2017.
- [5] Q. Ma and K. Tanaka, “ WebTelop: dynamic TV- content augmentation by using web pages ” . Proc. of IEEE International Conference on Multimedia and Expo (ICME 2003), Vol.2, pp.173-176, 2003.
- [6] TV TOKYO, “Let’s go somewhere. It is not only Kanazawa! The best Hokuriku Shinkansen Line stations you should get off,” April, 2015.
- [7] Yomiuri Telecasting Corporation,”Ten, Ichiban! The Charming Otona Trip! Sneak into the Popular Tour in Kobe,” September, 2017.
- [8] Yomiuri Telecasting Corporation,”Ten, Travel Concierge! Selected Gourmets of Gion, Kyoto,” September, 2017.
- [9] NHK Educational TV Osaka, ” Ability of the Ancestor, Wisdom Spring, The Third generation Iemitsu Tokugawa, How He Inherited the Composition Well,” September, 2017.