International Journal of Advanced Intelligence
Volume 10, Number 1, pp.113-119, March, 2018.

Index Generation of BGM Video Based on Distinctive Comments

Hirokimi Fukuda
Faculty of Engineering,Tokushima Univercity,
2-1 Jousanjima-cho,Tokushima city ,Tokusima,770-0814, Japan
c501737037@tokushima-u.ac.jp

Kazuyuki Matsumoto, Minoru Yoshida, Kenji Kita
Tokushima Univercity
matumoto@is.tokushima-u.ac.jp
mino@is.tokushima-u.ac.jp
kita@is.tokushima-u.ac.jp

Received (15 Sep. 2017)
Revised (15 Dec. 2017)

The purpose of this study is to generate indexes for background music (BGM) based on distinctive comments annotated to BGM videos posted on Nico Nico Videos. Our proposed method detects the end/start positions of the BGM considering the increase/decrease in the number of comments, comments distinctive to the songs, and exclamatory expressions annotated to the videos. The evaluation result indicated that the proposed method could generate correct indexes within a 10 s error for BGM videos less than 40 min long and with annotations of over 30,000 comments.

Keywords: Nico Nico Video; index generation; BGM video for work.

1. Introduction

The number of users of video sharing sites such as Nico Nico Video[1] and YouTube[2] has been increasing year by year. Tens of millions of these users are young people, and more than 30 million videos have currently been uploaded on Nico Nico Video. Among these,"BGM video for work" accounts for the largest number of videos, and are also viewed the most frequently. Moreover, videos under "BGM video for work" contain several songs.

However, most of the uploaded videos cannot be understood by simply glancing at their content. Accordingly, video viewers need to expend large amounts of time and considerable labor playing the entire video to check its content. These aspects often prevent users from viewing videos comfortably. Therefore, an automatic indexing system for videos would be beneficial. In this research, we propose a system that generates indexes of song titles and the starting positions for each song, by analyzing comment data posted for the videos.

2. Related research

Aoki et al.[3] proposed a method to detect the climactic part of music used for videos and thereby generate a summary of the videos based on the frequency of comments. Videos on Nico Nico containing funny scenes appear to have more comments, and these scenes should be included in video summaries. However, if other contexts are necessary to understand the humor in these scenes, Aoki et al. 's system cannot function suitably with regard to generating video summaries. Additionally, the method may not work well with cooking programs; sometimes, talking scenes are included in between the cooking processes, which might be preferentially selected over scenes containing cooking processes.

Nan et al.[4] proposed a data analysis system for comments on short online videos. Matsumoto et al.[5] devised a method to detect unfair videos using short comments posted for the videos. Takano et al.[6] used a real-time communication service for audiences watching a live video program. Takama et al.[7] proposed a parallel movie presentation for video summaries, which generates a video summary from the video content and its respective comments.

Other related approaches that exclusively use the frequency of comments suffers from limitations in that they cannot be applied to videos intended to collect as many comments as possible. In this research, we solve the above-mentioned problems by considering the content of the comments.

3. Proposed method

3.1. *Distinctive Comment*

Nico Nico videos possess a function that enables users to post comments with reference to specific scenes in a video. BGM video for work videos allow distinctive comments corresponding to a song to be posted each time the song is switched. In our study, we define such comments as "distinctive comments." Several sites, such as Nico Nico Encyclopedia, collect distinctive comments because many users copy and paste comments from those sites to post them as their own comments. Therefore, the same comments are often observed to be posted for certain songs in BGM video for work videos. By detecting these distinctive comments, we believe that it would be possible to determine whether certain songs appear in BGM videos for work. Table 1 shows the example of distinctive comments which were posted as the video comments.

3.2. *Video Splitting*

Video content immediately before a song begins tends to be lively, thereby attracting many comments. The majority of the comments for such scenes include exclamatory expressions such as "Uooooo" and "Kitaa—-."

First, we sort the comment data by reproduction time. Because the posting times of comments are recorded and indicated in units of 1/100th of a second, we round

Table 1. Example of distinctive comments

Song title	Distinctive comments	Romanized
Daydream cafe	心がぴょんぴょんする	*Kokoro ga pyon-pyon suru*
true my heart	きしめん	*Kishimen*
Rising Hope	ある意味ネギトロ	*Aruimi negitoro*
TILL THE END	課税のせい	*Kazei no sei*
dots and lines	土佐犬は出ません	*Tosaken wa demasen*
grilletto	たまねぎ干されているようで	*Tamanegi hosareteiru youde*

the values down to N seconds and count the number of comments per N seconds. We regard 00:00 seconds from the beginning of the video as the start position of the first song. We investigate how the number of comments changes from $(i-1)N$ seconds until iN seconds (i is a positive integer value).

By Equation 1, the system detects the points where the numbers of comments increase. Then, we empirically identify points where the number of comments increases when $R_x(i, N)$ is 2.5 or more. Term c_t indicates the number of comments at t.

Next, we analyze the comments posted at the above-mentioned detected points. The system matches sets of comments with the exclamation list, and excludes points with a low matching ratio using Equation 2. Term e_i indicates the number of exclamation comments at i when $R_x(i, N) \geq 2.5$. c_i indicates the number of comments at i.

In this study, we empirically exclude points if the song starts when R_y is less than 0.4. Finally, the points that meet all the above-mentioned criteria are judged to be the song starting points, and they are annotated with indexes.

$$R_x(i, N) = \frac{\sum_{t=iN}^{(i+1)N} c_t}{\sum_{t=(i-1)N}^{iN} c_t} \tag{1}$$

$$R_y(i) = \frac{e_i}{c_i} \tag{2}$$

4. Experiment and Discussions

4.1. *Outline of Experiment*

In this experiment, we target videos that are annotated with the BGM video for work tag on Nico Nico Video. We exclude videos for which the number of comments exceeds the replay frequency. This is because many of the comments are not related to the video content.

We conduct the experiment by categorizing the evaluation videos in the video set by the number of comments posted for each video. The result is evaluated by

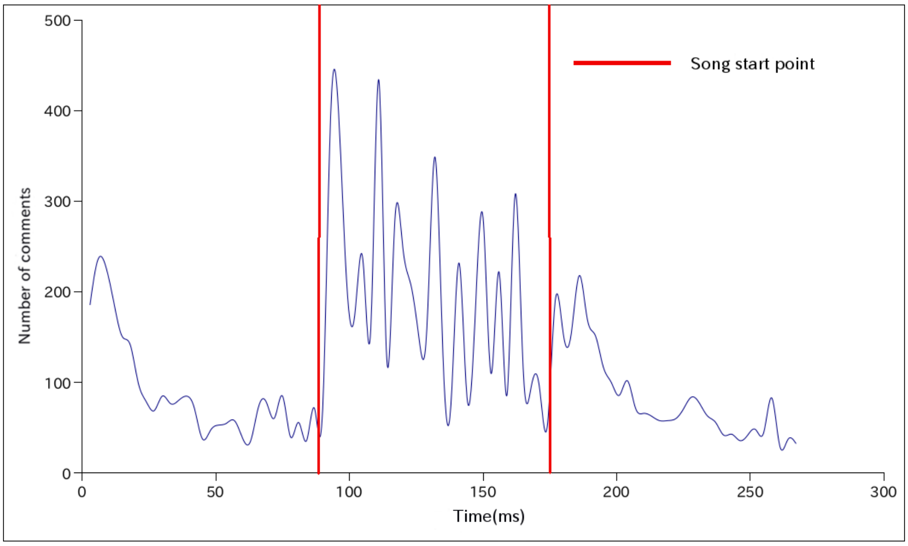

Fig. 1. Examples of increasing the number of comments at breaks of songs

using the song title and correct start position, which are annotated manually as the gold standard. Table 2 shows the number of comments and number of video for the experiment.

Table 2. Data used for experiments

Number of comments	Number of video
100 thousand or more	2
50,000 to 90,000	3
40,000	3
30,000	3
20,000	3
10,000	3
5,000	3
2,525	3

4.2. *Result*

Table 3 shows the summary of experimental results. In the column of "Maximum error(s)", the averaged errors of detected point are shown by seconds. And, in the column of "Estimate of song title(success rate)", the averaged success rates of song title estimation are shown.

Table 3. Summary of experimental results

Number of comments	Maximum error(s)	Estimate of song title(success rate %)
100 thousand or more	+14	23/45 (51.11)
50,000 to 90,000	+8	49/67 (73.13)
40,000	+18	55/71 (77.46)
30,000	+18	74/112 (66.07)
20,000	+144	71/108 (65.74)
10,000	+101	28/73 (38.35)
5,000	+121	22/65 (33.84)
2,525	+484	17/88 (19.31)

4.3. *Discussions*

One commonly observed tendency was that the judged starting points of the songs appeared backward from the correct positions. This error may be attributed to the timing of comment posting usually lagging behind the start time of the video being played, which is a consequence of the manner in which the commenting system of Nico Nico Videos is designed. Because users post their comments after watching certain video scenes, a gap of several second exists from the starting point.

Moreover, many comments are posted in the first half of the video, and the error range is small; however, the error range expands in the second half of the video. This tendency was noted more often in videos of longer duration. Conversely, videos using the ranking format exhibited higher accuracies in the latter half. This might be because famous songs appear in the higher ranks; as a result, comment postings increased, resulting in an increase in the accuracy.

When videos had less than 10,000 comments and their duration was longer than approximately 10 s, unlike "climax medley," the system could provide more accurate estimates of the starting points of songs. Climax medley refers to videos that summarize only the climax part of songs in a medley format. Moreover, song title estimation showed higher accuracy for videos that could not be regarded as climax medley and those with more than 10, 000 comments.

We could not obtain successful results for videos posted before 2010 because our study focused on the distinctive comments made for animation or game videos posted on or after 2010. Because many videos with numerous comments are old, and many years have passed since they were posted, the system could not provide accurate estimates of the starting points for such videos.

Some songs that were not included in the videos were estimated nonetheless, possibly because many distinctive comments for other songs were posted for these videos. Although distinctive comments appeared for videos with fewer comments, the system could not estimate the song titles because these comments were infrequently made. However, for videos with fewer comments, we could evaluate whether they estimated the correct song titles without considering the starting points of the songs. As a result, the system could successfully estimate approximately 80% of the song titles correctly.

5. Conclusion

In this research, we proposed a method to create an index for BGM videos by using their comment information. The proposed method allowed the creation of an index exclusively from comment information without using information on the video or its audio. Concretely, it was possible to create an index within a 10 s error range for videos of 40 min duration and with more than 30, 000 comments. In this experiment, we could not estimate the starting points of songs for videos with fewer comments, because when the number of comments is small the increase rate of the comments becomes low.

In future work, we would like to solve the above-mentioned problems by using semantic features, meaning features other than changes in the number of comments or time of comment posting.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 15K00425, 15K00309, 15K16077.

References

1. Nico Nico Video : `http://www.nicovideo.jp/`.
2. YouTube : `https://www.youtube.com/`
3. Hidenori Aoki and Homei Miyashita : A Trial for Video Summarization and Chorus-Section Detecting on Nicovideo, IPSJ SIG technical reports 2008(50(2008-HCI-128)), pp. 37–42, 2008-05-21.
4. Nan Yang and Shufan Zhang : Data Analysis System for Online Short Video Comments, Proc. of the 15th International Conference on Computer and Information Science (ICIS), 2016.
5. Kazuyuki Matsumoto, Hayato Shimizu, Minoru Yoshida and Kenji Kita: Time-Series Analysis of Video Comments on Social Media, Sociolinguistics – Interdisciplinary Perspectives, Chapter 7, pp.105–123, DOI: 10.5772/intechopen.68636
6. Yutaro Takano, Kohta Ohshima, Koji Tajima, Osamu Takada, Matsuaki Terada : A Real Time Communication Service for Audience in Video Program Simultaneous Watching, The IEICE transactions on information and systems 93(10), 2302-2316, 2010-10-01.
7. Yasufumi Takama , Duo YU : Proposal of Parallel Movie Presentation for Video Summary The IEICE transactions on information and systems, Vol. J98-D, No. 5, pp. 802–810, 2015.

Hirokimi Fukuda

He received bachelor's degree in 2017 at Tokushima University. His research interests include Natural Language Processing, Web Intelligence and Data Science.

Kazuyuki Matsumoto

He received the PhD degree in 2008 from Tokushima University. He is currently an assistant professor of Tokushima University. His research interests include Affective Computing, Emotion Recognition, Artificial Intelligence and Natural Language Processing. He is a member of IPSJ, ANLP, IEICE and IEEJ.

Minoru Yoshida

He is a lecturer at the Department of Information Science and Intelligent Systems, University of Tokushima. After receiving his BSc, MSc, and PhD degrees from the University of Tokyo in 1998, 2000, and 2003, respectively, he worked as an assistant professor at the Information Technology Center, University of Tokyo. His current research interests include Web Document Analysis and Text Mining for the Documents on the WWW.

Kenji Kita

He received the B.S. degree in mathematics and the PhD degree in electrical engineering, both from Waseda University, Tokyo, Japan, in 1981 and 1992, respectively. From 1983 to 1987, he worked for the Oki Electric Industry Co. Ltd., Tokyo, Japan. From 1987 to 1992, he was a researcher at ATR Interpreting Telephony Research Laboratories, Kyoto, Japan. Since 1992, he has been with Tokushima University, Tokushima, Japan, where he is currently a Professor at Faculty of Engineering. His current research interests include multimedia information retrieval, natural language processing, and speech recognition.