PAPER Special Section on Intelligent Information and Communication Technology and its Applications to Creative Activity Support

Detecting TV Program Highlight Scenes Using Twitter Data Classified by Twitter User Behavior and Evaluating It to Soccer Game TV Programs

Tessai HAYAMA[†a)], ***Member***

SUMMARY This paper presents a novel TV event detection method for automatically generating TV program digests by using Twitter data. Previous studies of TV program digest generation based on Twitter data have developed TV event detection methods that analyze the frequency time series of tweets that users made while watching a given TV program; however, in most of the previous studies, differences in how Twitter is used, e.g., sharing information versus conversing, have not been taken into consideration. Since these different types of Twitter data are lumped together into one category, it is difficult to detect highlight scenes of TV programs and correctly extract their content from the Twitter data. Therefore, this paper presents a highlight scene detection method to automatically generate TV program digests for TV programs based on Twitter data classified by Twitter user behavior. To confirm the effectiveness of the proposed method, experiments using 49 soccer game TV programs were conducted.

key words: *highlight-scene detection, TV digest generation, Twitter data, burst detection*

1. Introduction

This paper describes a novel TV-event detection method for automatically generating TV program digests using Twitter data. TV program digests are useful for understanding the overview of a given TV program and for quickly searching for interesting scenes in the TV program. They are therefore often used in news program, program propaganda, and so on; however, high costs in terms of time and labor are required to manually create digests for such TV programs. Here it is necessary to find and highlight scenes in the TV programs and then create indices for these highlighted scenes. To reduce costs, automatic video digest generation methods have been developed using the audiovisual features of the video [2], [14], [16]. Though such methods successfully extract well-defined scenes from videos with high accuracy, it is difficult for these methods to extract undefined highlight scenes; moreover, it is also difficult to assign indices that include various types of information, including keywords and impression words, to the extracted scenes.

In recent years, the number of users who use Twitter while watching TV programs has increased. Users discuss and share details and opinions of what happens in the given TV program on Twitter in real time, such that there are a huge number of live tweets regarding the TV programs* [9]. Previous studies have developed TV event detection methods for TV program digest generation based on live tweets corresponding to the given TV programs [6]–[8], [10], [11]. These methods analyze trends in the frequency time series of the tweets that users made while watching a given TV program. Since Twitter users behave in different ways, for example, conversing or sharing information, Twitter content is not always synchronized with the TV program content in real time. It is therefore difficult to correctly detect the highlight scenes of TV programs from the Twitter data.

Given the above problems, this paper proposes a highlight scene detection method for TV programs based on Twitter data categorized by Twitter user behavior with high accuracy [3], [4]. Using the categorized Twitter data, the proposed method detects highlight scenes of TV programs. The effectiveness of the proposed method has been confirmed by experiments conducted using 49 soccer game TV programs.

2. Related Studies

To automatically generate sports TV digests, some researchers have analyzed visual and audio features of sports TV program videos, for example, recognizing goal posts within soccer video scenes to detect goal scenes of a soccer game [14], developing an image/speech hybrid recognition method for detecting highlight scenes of soccer game videos, developing a Hidden Markov Model (HMM) method that assigns labels to highlight scenes of soccer game TV programs by using the speech features of the videos [16], and developing an emotion model constructed from visual and audio features of videos to detect highlight scenes of soccer game TV programs [2]. Although the above studies have been able to accurately detect the main scenes within well-defined visual and audio features, it has been impossible for them to detect undefined main scenes or their achievements have been prone to false positives despite high recall rates.

Also, in recent years, researchers have developed high-

Manuscript received March 15, 2017.
Manuscript revised July 22, 2017.
Manuscript publicized January 19, 2018.

[†]The author is with Department of Information & Management Systems Engineering, Nagaoka University of Technology, Nagaoka-shi, 940–2188 Japan.
a) E-mail: t-hayama@kjs.nagaokaut.ac.jp

DOI: 10.1587/transinf.2016IIP0020

*http://www.nielsen.com/us/en/press-room/2012/nielsen-and-twitter-establish-social-tv-rating.html

light scene detection methods for TV programs using Twitter data, based primarily on the analysis of the frequency time series of tweets that users made while watching a given TV program. The literature [8] describes the ability of such methods to detect highlight scenes of TV programs using Twitter data with accuracies similar to that of the aforementioned audio and visual analysis. Some approaches have analyzed Twitter data using categories of tweets and Twitter users to detect highlight scenes of TV programs based on a variety of viewpoints. As an example of tweet categorization, tweets may be categorized into groups based on such emotional representations as emotion words [1] and emoticons [15]; then the main scenes of the given TV program are differently detected on the basis of these categories. As an example of user categorization, Twitter users may be separated by the fun side of each team to detect highlight scenes from opposite viewpoints in such team games as soccer [12], baseball [6], and football [13]. Based on the tweets of one of these separated users, enjoyable and exciting scenes of the game can be detected for each fun side.

These previous studies have analyzed relationships between highlight scenes of a TV program and tweets made during the time of a given scene by observing the trends of frequency time series of tweets and adapting Twitter data with a burst detection method. Twitter data used in previous studies have been collected by hashtags and keywords; then all such data has been used for analysis. Since Twitter user behavior manifests in different ways, Twitter content is not always synchronized with the content of the TV program in real time. It is therefore difficult for the previous approaches to correctly detect highlight scenes of TV programs from Twitter data. As noted above, this paper introduces the categorization of user behavior into highlight scene detection. The approach is different from the previous studies in that tweet categorization is based on user behavior, not on linguistic representation.

3. Proposed Method

This paper proposes a novel method to use Twitter data to detect highlight scenes of TV programs. To achieve this, the method categorizes Twitter data into user types during a timeslot of the TV program, and then properly applies one of the categorized Twitter data to the detection method of the highlight scenes. As shown in Fig. 1, the procedure of the proposed method consists of three steps. In the first step, Twitter users generate tweets while watching the TV program; these tweets are categorized into groups based on user types during the given timeslot of the TV program. In the second step, the frequency time series of the tweets for each user type is created. In the final step, the burst points of the frequency time series are automatically detected to decide the time periods of the highlight scenes of the TV program.

In the given procedure, highlight scenes of the TV program is identified with a burst detection method using the Twitter data. If this approach properly adapts the tweets of certain user types to the burst detection method, the approach performs better than an approach that lumps user types into one category. For example, when a user simply posts tweets regarding a TV program on Twitter while watching the TV program, the user's tweets tend to sensitively respond to the main scences on the TV program in real-time because the tweets are often posted at once after each event. Conversely, when a user communicates his or her impression of the TV program to other users, the user's tweets are likely to be out of sync with the content of the TV program and likely not include enough information regarding the TV program, instead including brief feedback, off-topic conversations, and the like-though it is interesting to include words that represent the user's impression of the TV program. Thus, the proposed method introduces analysis of user behavior into the detection method of the highlight scenes of TV programs such that remarkable scenes are extracted with high accuracy from the categorized Twitter data.

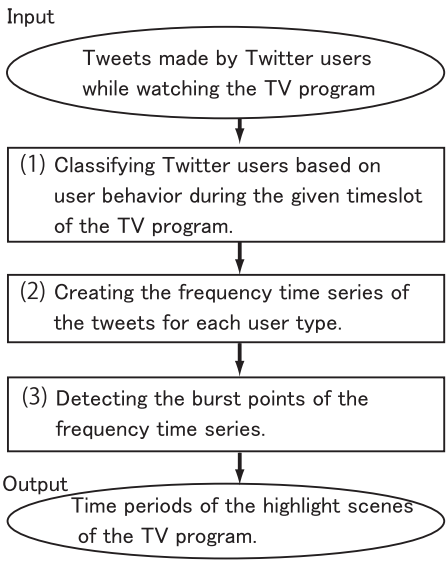

Fig. 1 Procedure of the proposed method.

Details of each step of the procedure are described below:

(1) Classifying Twitter users based on user behavior during a timeslot of the TV program.

Twitter users tweeting while watching the given TV program are classified into types of similar tweet behavior using cluster analysis. Considering how Twitter is used, the following four features were chosen for cluster analysis and investigated for an hour-long TV program for every user:

- The number of retweets and reply-tweets, which represents the degree of relationships between the given user and other users. If users often generate tweets that include retweets or reply-tweets, users place a priority on the Twitter conversation.
- The number of hashtags, which represents the degree of contribution of information sharing. The hashtag

of a topic is useful for retrieving tweets of the given topic. If users often generate tweets that include hashtags with regard to the TV program, users put a priority on information sharing among Twitter users who are interested in the same TV program.
- The average number of characters, which represents the degree of informative content. If users often generate tweets that include large numbers of characters, the tweets are likely to include large numbers of keywords regarding the given TV program.
- The average number of tweets, which represents the sensitivity in regards to the TV program and other tweets. If users make a large number of tweets during the timeslot of the TV program, they tend to sensitively respond to the main scenes of the TV program.

In my previous study [3], to detect highlight events and generate metadata to the events from twitter data of a soccer game, the twitter data was categorized into four user groups of which each had different behaviors on Twitter as follows; "heavy use of hashtags," "heavy use of retweets", "parallel use of retweets/plain-tweets," and "heavy use of plain-tweets." Therefore, the system adopted classifying the Twitter data into the four user groups by using the Ward's method [17] because of using the same features and the clustering method as the previous research. The Ward's method is used to calculate linkage for aggregative hierarchical clustering. In the method, all clusters are then compared, and the pair of clusters with the smallest distance between them is selected based on the squared Euclidean distance as Eq. (1) and merged into a single cluster. The above two steps are repeated until there is specified number of clusters.

$$Dis(c_i, c_j) = \sqrt{\sum_{k=1}^{N}(c_{ik} - c_{jk})} \tag{1}$$

Here, $Dis(c_i, c_j)$ represents distance value between the ith cluster and the jth cluster. C_{ik} represents the kth feature value within the ith cluster for the cluster analysis.

(2) Creating the frequency time series of the tweets for each user type.

To create the frequency time series of the tweets for each user type, tweet frequencies for certain a time span are calculated and arranged into the timeline of the TV program. The time span of the current system is set to 5 s, which was determined by investigating the time gaps between highlight scenes of the TV program and the tweets regarding these scenes [3].

(3) Detecting highlight scenes of the TV program from the frequency time series of the tweets via the burst detection method.

To detect highlight scenes of the TV program, the unusual increases in the frequency time series of the tweets must be found. The current system adopts Kleinberg's method [5] for burst detection; this method is based on a HMM algorithm that automatically detects abnormal time spans and their burst level on time-series data. The burst detection procedure consists of the following three steps:

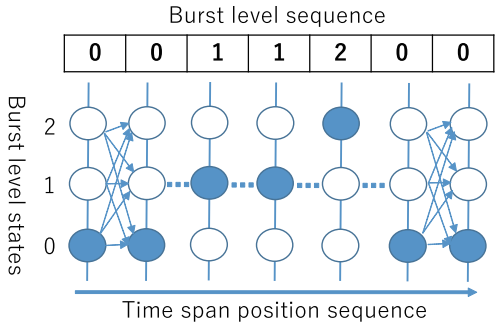

Fig. 2 Burst detection algorithm based on hidden Markov mode.

(a) An emergence possibility for each burst-level state of every time span position is calculated based on the tweet frequency of the position. The emergence possibilities of all the state as shown in Fig. 2 are computed by Eq. (2).

$$f_{it}(x_i) = \begin{cases} \alpha_0 = \dfrac{n}{T} & (i = 0) \\ \alpha_i = \dfrac{n}{T} s^i & (i > 0). \end{cases} \tag{2}$$

Here, x_i, $f_{it}(x_i)$, n, T, and s represent the ith burst level state, an emergent possibility for x_i on position sequence of time t, number of tweets in a timeslot of the given TV program, time length of the given TV program, and a scaling parameter ($s > 0$), respectively.

(b) All time sequences of the burst-level states are found and their costs are calculated based on the emergence possibilities included in the time sequences by Eq. (3).

$$C(q|x) = \left(\sum_{t=0}^{n-1} \Gamma(i_t, i_{t+1})\right) + \left(\sum_{t=1}^{n} -\ln f_{it}(x_t)\right), \tag{3}$$

$$\Gamma(i_t, i_{t+1}) = \begin{cases} 1 & (i_{t+1} > i_t) \\ 0 & (i_{t+1} \le i_t). \end{cases} \tag{4}$$

$C(q|x)$ represents cost for time-position sequence $x = \{x_1, \ldots, x_n\}$ and state sequence $q = \{q_{i1}, \ldots, q_{in}\}$ during a timeslot of the given TV program. $\Gamma(i, j)$ represents cost associated with a state transition from q_i to q_j, meaning that the cost is added only when moving from a lower-intensity burst state to a higher-intensity one.

(c) The time sequence of the burst-level with the minimum cost is chosen from all time sequences, and the time periods with the burst in the sequences are determined as the highlight scenes.

In Fig. 2, when the time sequence composed by filled circles is chosen as the one with the minimum cost, the burst level sequence to the time span sequence with the burst level is obtained as {0, 0, 1, 1, 2, 0}.

Note that the system sets 3 as upper limit value of burst

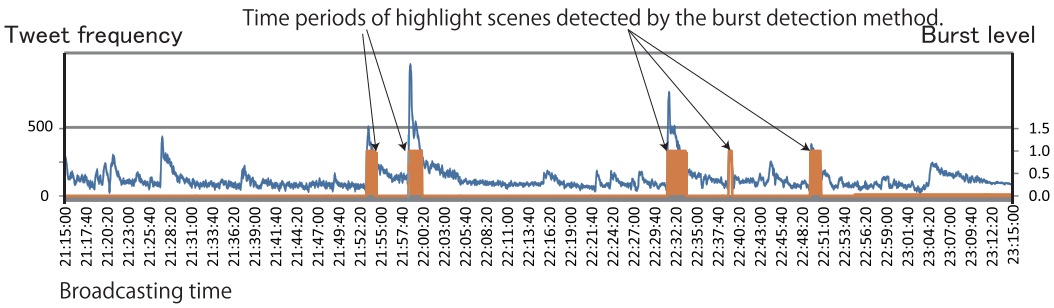

Fig. 3 An example of adapting the burst detection method to the frequency time series of tweets.

Table 1 Twitter data that was used in the experiment.

	All tweets		Tweets with hashtag		Retweets/Reply-tweets		Plain-tweets	
	tweet	user	tweet	user	tweet	user	tweet	user
Number of games	49		49		49		49	
Total number	6770359	1068245	628048	176189	2469413	583647	3948565	811043
Average number	138170.59	21800.92	12817.31	3595.69	50396.18	11911.16	80582.96	16551.90
(Min. per game)	26388	6564	3568	1157	8817	2980	15266	4575
(Max. per game)	1051716	85178	95880	18459	501900	61345	494845	66078
(SD.)	24805.43	2075.23	2087.10	444.79	10417.43	1388.62	14023.21	1749.61

level i and adopts the Viterbi Algorithm to choose the sequence with the minimum cost.

An example of adapting the burst detection method to the frequency time series of tweets is shown in Fig. 3.

4. Experiment

4.1 Overview

This paper proposes a method to detect remarkable scenes with high accuracy using Twitter data categorized by user behavior. In the experiment, the proposed method was compared to a burst detection method that adopts noncategorized Twitter data; the focus of the experiment was the accuracy of highlight scene detection of soccer game TV programs.

For the experimental data, 49 soccer game TV programs of 2014 FIFA World Cup from June 12, 2014 to July 13, 2014 were used, as was Twitter data consisting of 6,770,359 tweets collected from 1,068,245 of cumulative total Twitter followers of soccer player accounts. The details of the Twitter data is shown in Table 1.

The dataset for the experiment was generated from live news feeds of the soccer game on the Web[†], which reported the main events, such as shots, goals, fouls, and substitutes, as well as their times. Each the event and its time of the soccer games were manually confirmed with the TV program videos of the soccer games.

4.2 Results

(1) Classifying Twitter users based on user behavior

The results of classifying Twitter users based on user behavior are shown in Fig. 4 and Table 2.

[†]http://www.nikkansports.com/brazil2014/

- Group 1, which is labeled as "heavy use of plain-tweet." The users of this group generated tweets that included much higher rate of the plain-tweet use (i.e., 0.78) than the ones of the retweets/reply-tweets use (i.e., 0.21) and the hashtag use (i.e., 0.02). Number of users of the group (i.e., 9309.69) was the largest of the four groups and average number of the tweet characters (i.e., 19.12) was the smallest of the four.
- Group 2, which is labeled as "heavy use of retweet/reply-tweet." The users of the group generated tweets that included much higher rate of the retweet/reply-tweet use (i.e., 0.60) than the ones of plain-tweet use (i.e., 0.37) and hashtag use (i.e., 0.20). The rate of the hashtag use was the largest among all the groups. Number of users of the group (i.e. 2240.12) and average number of the tweets per game (i.e., 3.02) were the smallest of the four. On the other hand, number of tweet characters of the group (i.e., 126.33) was the largest of the four.
- Group 3, which is labeled as "parallel use of plain-tweet and retweet/reply-tweet (more plain-tweet use)." The rates of the plain-tweet use (i.e., 0.53) and retweet/reply-tweet use (i.e., 0.42) of the group were in middle, and the rate of the plain-tweet use of the group was higher than the one of the retweet/replay-tweet use.
- Group 4, which is labeled as "parallel use of plain-tweet and retweet/reply-tweet (equal use)." The user behavior of the group while watching a TV program is the similar to the one of the group 3, however there was not statistically significant difference between the rates of the plain-tweet use (i.e., 0.47) and the retweet/reply-tweet use (i.e., 0.47). Number of tweet characters of the group (i.e., 65.28) was larger than the one of the group 3 (i.e., 55.09).

The categorized user group 3 and 4 have very similar

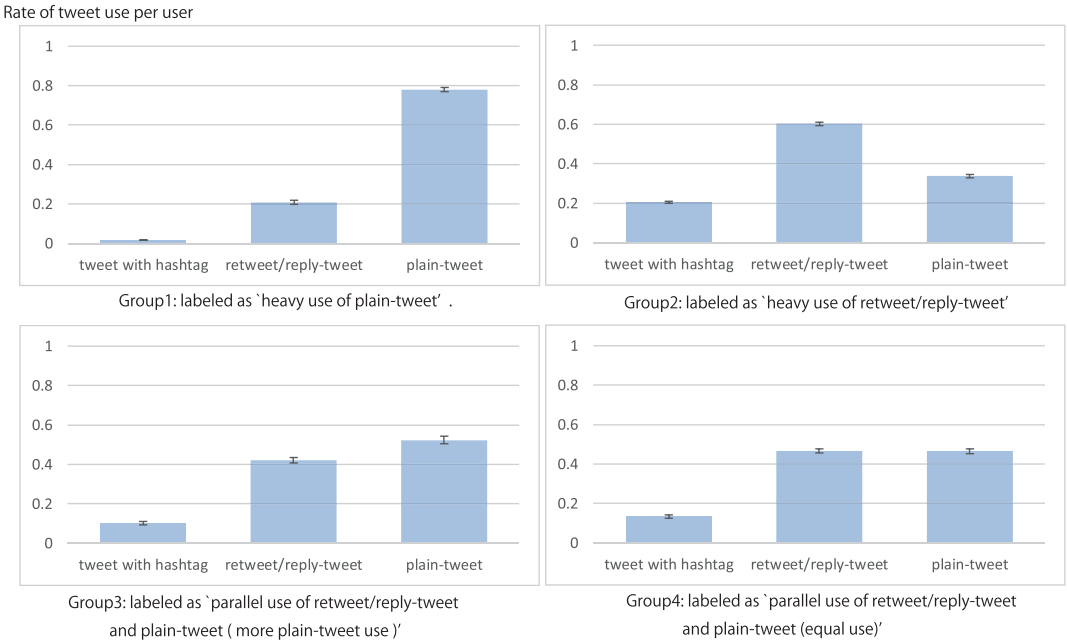

Fig. 4 Results of classifying Twitter users based on user behavior types.

Table 2 Results of classifying Twitter users based on user behavior types.

	Average number of users		Average number of tweets		Average number of tweet characters	
Group 1	9309.69	(946.69)	5.20	(0.26)	19.12	(0.27)
Group 2	2240.12	(139.47)	3.02	(0.12)	126.33	(0.37)
Group 3	6078.43	(625.49)	6.12	(0.30)	55.09	(3.10)
Group 4	4544.35	(551.65)	5.53	(0.23)	65.28	(2.31)

() shows SD. value for each item.

features but difference on whether there is significant difference between the use rates of plain-tweet and retweet/reply-tweet, so that the properties within the categorized user groups are different to the ones within the categorized user groups in my previous work [3]. However, both of the user group 3 and 4 were adopted to investigate detecting the highlight-scenes in the experiment since the balance of using plain-tweet and retweet/reply-tweet affected how many keywords which represent the event contents were included in the tweets on the detected time periods according to the previous work. Therefore, Twitter data categorized into the above groups was adopted for the highlight scene detection in the experiment.

(2) Detecting highlight scenes of the TV program from the frequency time series of tweets based on the classification of user behavior

The accuracy and time of detecting the highlight scenes of soccer game TV programs from the frequency time series of tweets were investigated in the experiment. Table 3 shows the accuracy of detecting highligh scenes of the 49 soccer game TV programs. The recall and precision shown in Table 3 were calculated by the following measures; the recall is the ratio of the number of the detected events to the total number of the events in all the games and the precision is the ratio of the number of the correctly detected events to the number of the detected events.

In the precision results of detecting the highlight scenes, the highlight scenes were detected at high rate (i.e., over 0.91) in all cases of using one of the compared Twitter data. Using the categorized Twitter data labeled as "heavy use of plain-tweet," provides the highest precision of the highlight scene detection among them (i.e., 0.97). In the recall accuracy of detecting the highlight scenes, comparing to the five kinds of Twitter data with the types of the event scenes, all the types of the event scenes were detected at the highest rates (i.e., 0.94 for goal, 0.11 for shot, 0.16 for foul, and 0.10 for substitute) using Twitter data of Group 1, meaning that using one of the categorized Twitter data, labeled as "heavy use of plain-tweet," provided higher recall accuracy of the highlight scene detection than the rates (i.e., 0.82 for goal, 0.07 for shot, 0.11 for foul, and 0.06 for substitute) of using the noncategorized Twitter data. On the other hand, using one of the categorized Twitter data, labeled as "heavy use of retweet/reply-tweet," provided much lower recall accuracy of the detection than the other methods. In each the event-type which was detected by using Twitter data of Group 1, although the goal scenes were detected at high rate, the scenes of shot/foul/substitute were detected at low rate, i.e., average number of shot scenes was 1 per game and average numbers of foul and substitute scenes were 1 per 2 games, respectively. Generally, most of

Table 3 Accuracy of detecting highlight scenes of the 49 soccer game TV programs from the frequency time series of tweets.

Twitter data	Precision	Recall							
		Highlight-event types							
		Goal		Shot		Foul		Substitute	
Noncategorized data	0.96(206/215)	0.82	(102/124)	0.07	(75/1112)	0.11	(15/134)	0.06	(14/254)
Group 1: "heavy use of plain-tweet"	0.97(287/296)	0.94	(116/124)	0.11	(124/1112)	0.16	(22/134)	0.10	(25/254)
Group 2: "heavy use of retweet/reply-tweet")	0.92(11/12)	0.06	(7/124)	0.00	(1/1112)	0.01	(2/134)	0.00	(1/254)
Group 3: "parallel use of plain-tweet and retweet/ reply-tweet (more plain-tweet use)"	0.95(112/118)	0.48	(59/124)	0.03	(38/1112)	0.07	(9/134)	0.02	(6/254)
Group 4: "parallel use of plain-tweet and retweet/ reply-tweet (equal use)"	0.96(91/95)	0.40	(49/124)	0.02	(23/1112)	0.09	(12/134)	0.03	(7/254)

Table 4 Mean time of detecting highlight scenes of the 44 soccer-game TV programs by using noncategorized Twitter data and Twitter data of Group 1.

	Noncategorized Twitter data		Twitter data of Group 1	
Mean time per game to the detected events (sec.)	528.44	(SD. 30.30)	611.78	(SD. 24.98)
Mean time per game to the common detected events of both the data (sec.)	495.91	(SD. 31.48)	484.09	(SD. 24.85)

goal scenes are likely to be picked out as a highlight scene of soccer game; however, all the scenes of shot/foul/substitute are not picked out as the highlight scene. Therefore, the proposed method of using Twitter data of Group 1 detected event scenes of the soccer game as an appropriate highlight scene of soccer game.

Since plain-tweet takes up high percentage in the categorized Twitter data, labeled as "heavy use of plain-tweet," the recall accuracy of detecting the highlight scenes in the case of using only plain-tweet was evaluated. In the evaluation results, the method of using only plain-tweet detected the highlight scenes with the high recall accuracy (i.e., 0.94<116/124> for goal, 0.11<113/1112> for shot, 0.16<22/134> for foul, and 0.10<25/254> for substitute), meaning that, comparing to detecting the highlight scenes of using only plain-tweet, the detection method of using the categorized Twitter data detected more shot scenes and the same number of the other senses. Half of the games in which more shot scenes were detected by the proposed method were ended in a scoreless draw.

Next, time of detecting the highlight scenes was compared between using Twitter data of Group 1 and noncategorized Twitter data. The highlight-scene detection is often applied to summarize TV program movies and confirm the search results for the TV program movies. For the applications, the method should not only detect the highlight scenes with high accuracy but also extract the highlight scenes with as little waste of time as possible. Therefore, I evaluated the detection time of the proposed method. Table 4 shows the mean time of detecting highlight scenes of the 44 soccer game TV programs, excluding goalless games, by using the two data.

The mean times per game of the detected events in using Twitter data of Group 1 and noncategorized Twitter data were 528.44 and 611.78 seconds, respectively; there was not statistically significant difference between the two (i.e., $F(1, 86) = 2.51$, $p > .10$). The mean times per game of the common events detected from both the data were 495.91 and 481.09 in using the Twitter data of Group 1 and the noncategorized Twitter data, respectively; there was not statistically significant difference between the two (i.e., $F(1, 86) = 0.04$, $p > .10$). Therefore, although using Twitter data of Group 1 detected more highlight scenes of the TV programs than using the noncategorized Twitter data, the mean time per game of the detected scenes was not difference between the two data.

The 9 following events except the highlight event as shown in Table 4 were detected in the experiment; 4 offside goals, 3 brilliant plays of famous players, 1 spectator break-in event, and 1 appearing newsflash caption. The 8 events except appearing the caption can be treated as a highlight scene of soccer game.

4.3 Discussion

Through this experiment of using the 49 soccer game TV programs, the effectiveness of the proposed method was confirmed as follows:

- Twitter user behavior while watching the given TV programs was categorized into four types, i.e., "heavy use of plain-tweet," "heavy use of retweet/reply-tweet," "parallel use of plain-tweet and retweet/reply-tweet (more plain-tweet use)," and "parallel use of plain-tweet and retweet/reply-tweet (equal use)."
- It is useful to detect highlight scenes of TV program of soccer games by using tweets generated by users of the type "heavy use of plain tweets."
- When Twitter data categorized in "heavy use of plain tweets" was adopted to the proposed method, the proposed method detected appropriate highlight scenes of soccer games; most of goal scenes, some shot scenes, a few of scenes of foul/substitute, and so on.
- Although number of the highlight scenes in using Twitter data categorized in the "heavy use of plain-tweet" was larger than the one in using the noncategorized Twitter data, there was not statistically significant different time-length of the detected highlight scenes between the two data.

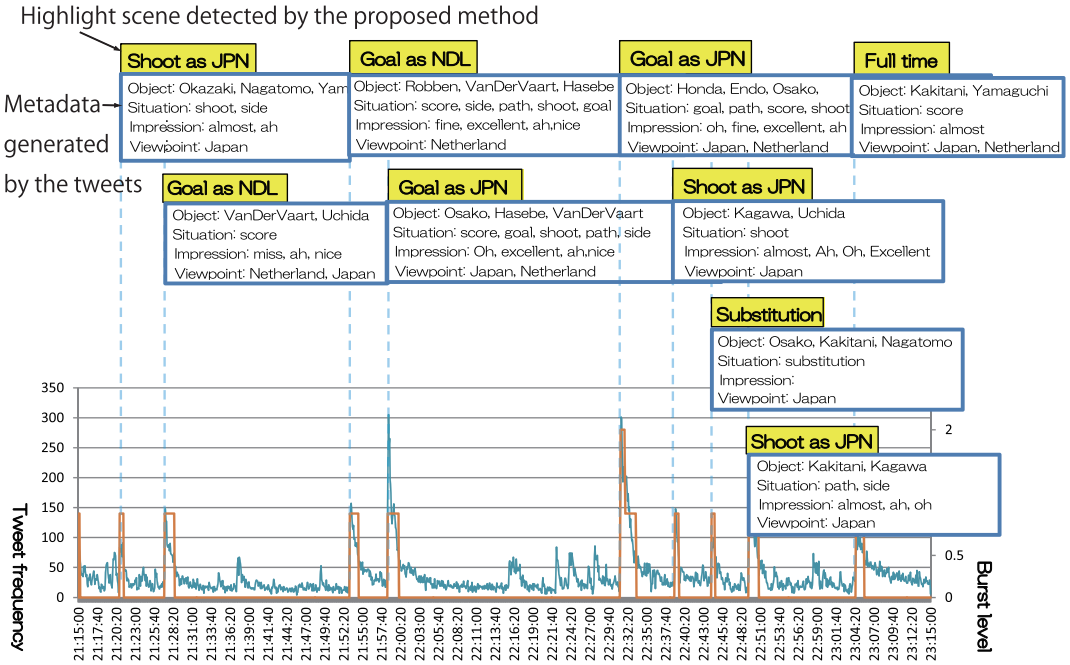

Fig. 5 An example of applying the proposed method to the highlighted scene detection of a succer game TV program.

Therefore, it is useful to introduce tweet categorization based on user behavior into detecting highlight scenes of a soccer game TV program.

Next, an example of applying the proposed method to the given soccer game TV program is shown in Fig. 5. For the example data, an international soccer game (i.e., Japan vs. Netherlands) broadcast in Japan on November 16, 2013 was used, as was Twitter data consisting of 376,656 tweets collected from 51,565 Twitter followers of soccer player accounts. Metadata attached to each detected scene was used bag-of-words extracted from the tweets. To extract the bag-of-words of each detected scene, the words included in the tweets were investigated using the burst detection method; then the unusual increasing words in the time window of the detected scene were determined as the metadata of the scene.

In the result, the content of the TV program is offered by the detected highlighted scenes and the extracted metadata. Thus, the result implies that metadata of bag-of-words to the highlight scenes is available via Twitter data.

5. Conclusions

In this paper, a novel method for detecting highlight scenes of a TV program is proposed. Previous related studies have dealt with Twitter data by lumping such data into one category and then detecting the highlight scenes of a TV program. The approach proposed in this paper introduced the analysis of user behavior into the method used to detect highlight scenes of TV programs. More specifically, the proposed method properly adapts tweets of one user type for correctly detecting highlight scenes. By conducting an experiment using 49 TV soccer programs, the effectiveness of the proposed method was confirmed.

In future work, a method which extracts metadata to each highlight scene of a given TV program from Twitter data will be developed. And also the proposed method will be applied to various TV programs, thus investigating the possibility of detecting highlight scenes of these programs and extracting the corresponding metadata. Further, a digest generation method for TV programs will be developed based on various kinds of viewpoints using the results the proposed method generates.

References

[1] K. Doman, T. Tomita, I. Ide, D. Deguchi, and H. Murase, "Event detection based on Twitter enthusiasm degree for generating a sports highlight video," Proc. 22nd ACM International Conference on Multimedia (ACM-MM2014), pp.949–952, 2014.
[2] A. Hanjali, "Adaptive extraction of highlights from a sport video based on excitement modeling," IEEE Trans. Multimed., vol.7, no.6, pp.1114–1122, 2005.
[3] T. Hayama, "Detecting TV program highlight scenes using Twitter data classified by Twitter user behavior," Proc. 10th International Conference on Knowledge, Information, and Creativity Support Systems, pp.164–175, 2015.
[4] T. Hayama, "Evaluating TV program highlight scene detection using Twitter data classified by Twitter user behavior and evaluating it to soccer game TV programs," Proc. 18th International Symposium on Advanced Intelligent Systems, pp.175–182, 2017.
[5] J. Kleinberg, "Bursty and hierarchical structure in streams," Proc. 8th ACM International Conference on Knowledge Discovery and Data Mining, pp.91–101, 2002.
[6] T. Kobayashi, T. Takahashi, D. Deguchi, I. Ide, and H. Murase, "Detection of biased broadcast sports video highlights by attribute-based Tweets analysis," S. Li et al. (eds.), Advances in Multimedia Mod-

eling, MMM 2013, Lecture Notes in Computer Science, vol.7733, pp.364–373, Springer, Berlin, Heidelberg, 2013.
[7] M. Kubo, R. Sasano, H. Takamura, and M. Okumura, "Generating live sports updates from Twitter by finding good reporters," Proc. 2013 IEEE/WIC/ACM International Conference on Web Intelligence, 2013.
[8] J. Lanagan and A.F. Smeation, "Using twitter to detect and tag important events in live sports," Proc. Fifth International Conference on Weblogs and Social Media, pp.542–545, 2011.
[9] Y. Mizunuma, S. Yamamoto, Y. Yamaguchi, A. Ikeuchi, T. Satoh, and S. Shimada, "Twitter bursts: Analysis of their occurrences and classifications," Proc. 8th International Conference on Digital Society 2014, pp.182–187, 2014.
[10] T. Nakahara and Y. Hamuro, "Detecting topics from Twitter posts during TV program viewing," Proc. IEEE 13th International Conference on Data Mining Workshops, pp.714–719, 2013.
[11] M. Nakazawa, M. Erdmann, K. Hoashi, and C. Ono, "Social indexing of TV programs: Detection and labeling of significant TV scenes by Twitter analysis," Proc. 26th International Conference on Advanced Information Networking and Applications Workshops, pp.141–146, 2012.
[12] G. van Oorschot, M. van Erp, and C. Dijkshoorn, "Automatic extraction of soccer game events from Twitter," Proc. Workhop on Detection, Representation, and Exploitation of Events in the Semantic Web, DeRiVE 2012, pp.21–30, 2012.
[13] A. Tang and S. Boring, "#EpicPlay: Crowd-sourcing sports video highlights," Proc. 30th Annual CHI Conference on Human Factors in Computing Systems, pp.1569–1572, 2012.
[14] T. Yamamoto, D. Shimizu, and M. Watanabe, "Research of automatic scene analysis for soccer broadcast images," Technical Report of the Institute of Electronics, Information and Communication Engineers, PRMU 104(573), pp.73–78, 2005 (in Japanese).
[15] T. Yamauchi, Y. Hayashi, and Y.I. Nakano, "Searching emotional scenes in TV programs based on twitter emotion analysis," Proc. 5th International Conference on Online Communities and Social Computing, pp.21–26, 2013.
[16] J. Wang, C. Xu, E. Chng, and Q. Tian, "Sport highlight dectection from keyword sequences using HMM," Proc. IEEE International Conference on Multimedia and Expo (ICME) 2014, pp.599–602, 2004.
[17] J.H. Ward Jr., "Hierarchical grouping to optimize an objective function," Journal of the American Statistical Association, vol.58, no.301, pp.236–244, 1963.

Tessai Hayama received his B.E. degree in Knowledge Engineering from Doshisha University in 2001, and M.E. and Ph.D. degrees in Knowledge Science from Japan Advanced Institute of Science and Technology in 2003 and 2006, respectively. From 2006 to 2012, he was an Assistant Professor with Knowledge Science of Japan Advanced Institute of Science and Technology. From 2012 to 2015, he was an Assistant Professor with Information Engineering of Kanazawa Institute of Technology. Form 2015 to 2016, he was an Associate Professor with Information Engineering of Kanazawa Institute of Technology. Currently he is an Associate Professor with Information and Management Systems Engineering of Nagaoka University of Technology. His research interests include knowledge systems, creative support systems and human interface.