Generation of a Video Summary on a News Topic Based on SNS Responses to News Stories

Kosuke Kato*
Nagoya University, Graduate School of Information Science
Furo-cho, Chikusa-ku, Nagoya, 464-8601 Japan
katok@murase.m.is.nagoya-u.ac.jp

Ichiro Ide
Nagoya University, Graduate School of Information Science
Furo-cho, Chikusa-ku, Nagoya, 464-8601 Japan
ide@is.nagoya-u.ac.jp

Daisuke Deguchi
Nagoya University, Information Strategy Office
Furo-cho, Chikusa-ku, Nagoya, 464-8601 Japan
ddeguchi@nagoya-u.jp

Hiroshi Murase
Nagoya University, Graduate School of Information Science
Furo-cho, Chikusa-ku, Nagoya, 464-8601 Japan
murase@is.nagoya-u.ac.jp

ABSTRACT

Archiving news videos is important since they accumulate valuable real-world information. When exploiting them, it is important to track the flow of news topics to understand them thoroughly. In order to do so, a method that structures the chronological semantic relations between news stories, namely the "topic thread structure" has been proposed in the past. However, simply viewing videos that compose this structure imposes a user to spend a long time watching detailed reports. On the other hand, Social Networking Services (SNS) have become very popular. SNS users often send and receive information in which they are interested while watching TV. Thus, we propose a method that automatically generates a video summary on a news topic from the general users' viewpoint based on responses of SNS users.

Categories and Subject Descriptors

H.3.1 [**Information Systems Applications**]: Abstracting methods

General Terms

Algorithms

Keywords

News video; SNS; topic thread structure; summarization

*Currently at Nihon Unisys, Ltd.

CrowdMM'15, October 30, 2015, Brisbane, Australia.

DOI: http://dx.doi.org/10.1145/2810188.2810189.

1. INTRODUCTION

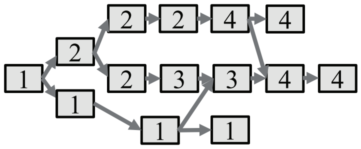
1.1 Overview

Recently, the diffusion of large capacity of storage has enabled us to easily store a large amount of broadcast videos in an online archive. Among various genres of broadcast videos, news videos are considered as particularly important since they accumulate rich information on events in the real world. When we consider news videos archived during a long period, it becomes important to track the flow of news topics comprehensively to understand them thoroughly, which is infeasible to do manually.

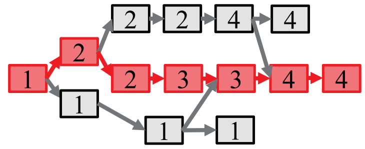
Therefore, works on the analysis of the semantics of news videos and the construction of a topic structure have been researched in the past. Most traditional approaches simply cluster the stories according to their contents (topic clustering) and then connect them linearly in chronological order. For example, Duygulu et al. [2] proposed a method that linearly connects news stories related to a specified news topic in chronological order. Wu et al. [3] proposed a method that constructs a binary tree corresponding to the change of the subject and context of news stories in a news topic cluster. However, these methods could not represent parallel multiple flows of a topic structure.

A more complex structure that can handle the transition of multiple topics can be represented as a directed tree, by considering both the chronological and the semantic relations between stories. However, a tree structure could not compactly represent the divergence and the reunion of the topic transition. So, Ide et al. [4] proposed a method that represents semantic relations between news stories as an acyclic directed graph called a "topic thread structure", which could represent parallel flows in a topic structure compactly. In this paper, we make use of this structure.

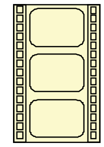
However, even with this kind of topic structure, it is time-consuming to actually view all the related videos. According to Ide et al. [4], it takes an average of 73 min. to view all the videos in a topic thread structure within an archive that



(a) Topic thread structure with topic clusters indicated by their IDs



(b) Representative topic thread



(c) Video summary

Figure 1: Approach for the generation of a video summary on a news topic.

consists of six months of news videos (Fig. 1 (a)). This could be reduced to an average of 46 min., if one topic thread is selected from the topic thread structure (Fig. 1 (b)). In order to select one representative topic thread from a topic thread structure automatically, Kato et al. [5] proposed a method that first estimates the representative ending news story, and next estimates the most representative topic thread. However, in this method, the representativeness was evaluated only from objective features obtained from the topic thread structure and the news stories, but not from the viewers' viewpoint. Thus, the authors claimed that the selected topic thread may not match the viewers' interest.

On the other hand, following the spread of various Social Networking Services (SNS), it has become common to post responses to them while watching TV in the so-called "second screen" fashion. Although this phenomenon is observed typically in sports contents, news contents also attract responses from SNS users.

So, in this paper, we first propose a method that automatically selects a representative topic thread (i.e. a path or a series of news stories) in a topic thread structure according to responses of SNS users (Fig. 1 (b)). We next propose a method that generates a video summary by extracting video segments corresponding to important sentences in the news story series that compose the selected topic thread, in order to further reduce the viewing time (Fig. 1 (c)). The importance of a sentence is again, defined according to responses of SNS users. In this way, the generated video summary should become more subjective towards the general viewers' viewpoint compared to traditional video summarization methods that generated relatively objective summaries. Although the viewers are not directly asked to provide responses to news stories for the purpose of video summarization and/or structuring as in a general crowdsourcing scheme, for example, taken by Smits and Hanjalic [6] and Snoek et al. [7], we consider the proposed approach that makes use of the crowd's, so to speak, "unconcious" activities, as an indirect source for crowdsourcing.

The paper is organized as follows. After briefly introducing the terminology and the topic thread structure in this section, in Section 2, we describe the proposed method. In Section 3, we report the results of an evaluation experiment. Finally, we conclude the paper in Section 4.

1.2 Definition of terms

Terms related to news contents that are used in this paper are defined as follows. First, the definition of basic terms by the TDT (Topic Detection and Tracking) Workshop organized by the National Institute of Standard and Technologies (NIST) is defined as follows [1].

- Event: Some incident that occurred at some specific time and place along with all necessary preconditions and unavoidable consequences.
- Story: A topically cohesive segment of news that includes two or more declarative independent clauses about a single event.
- Topic: A seminal event or activity, along with all directly related events and activities.

In addition, the following concepts and terms defined by Ide et al. [4] are as follows.

- Topic thread: A sequence of related stories chained chronologically. It may contain multiple topics.
- Topic thread structure: A directed graph composed of topic threads originating from a specified story.
- Topic cluster: A topically cohesive set of neighboring stories in the topic thread structure.

1.3 Topic thread structure and topic clusters

In this paper, we make use of the "topic thread structure" proposed by Ide et al. [4]. It is constructed by linking related news stories along the timeline starting from a specified news story, considering both semantic and chronological relations between news stories. It represents multiple flows (i.e. topic threads) of news topics as a directed acyclic graph.

Different from most other topic structuring methods, where related news stories are first clustered and then structured, this method first creates the topic thread structure which may contain multiple topics, and then extract topic clusters within the structure later. Figure 1 (a) is an example of a topic thread structure with IDs of topic clusters shown in each node (i.e. news story).

Since it is not the main interest of the work presented in this paper, we will refrain from introducing its algorithms in detail, and consider that the topic thread structures and the clusters within each of them are given and reasonable. Thus, in the experiment, only topic thread structures that made sense by the authors were included in the dataset.

2. GENERATION OF A VIDEO SUMMARY ON A NEWS TOPIC

2.1 Approach

We solve the task of generating a video summary according to responses of SNS users by the process flow shown in Fig. 2. We expect that a topic thread structure originating from a specified news story, and topic clusters in the

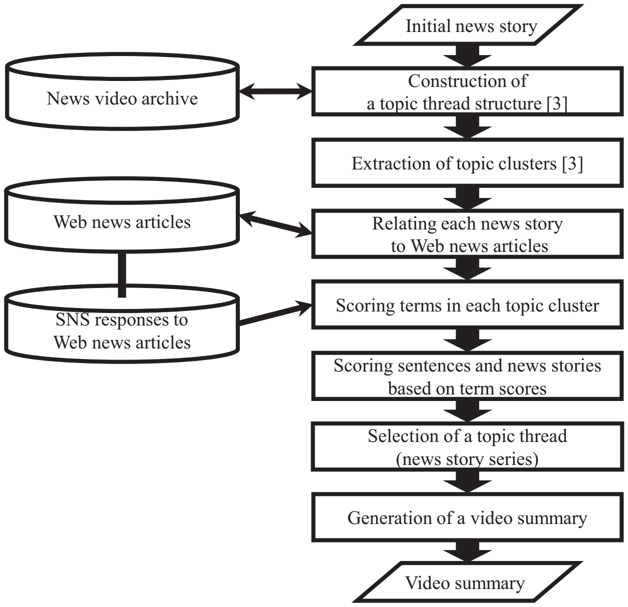


Figure 2: Process flow of the proposed method.

topic thread structure (Fig. 1 (a)) are given as input by the method explained in Section 1.3.

First, each news story in the topic thread structure are linked to Web news articles. Next, based on the number of SNS responses to each linked news article, each term that appears in a topic cluster is assigned a score. Then, based on the assigned term scores, sentences and subsequently, news stories, are scored. Finally, based on the story scores, a representative topic thread is selected from the topic thread structure (Fig. 1 (b)). Videos from the news stories that compose the selected topic thread are used for generating the video summary. Concretely, video segments corresponding to sentences with high scores are cropped and concatenated. Thus, we obtain a video summary on a news topic from the general users' viewpoint (Fig. 1 (c)).

2.2 Relating each news story to Web news articles

It is difficult to directly relate general responses of SNS users to a specific news story. So, we take an indirect approach; first, Web news articles are related to a news story, then SNS responses to the Web news articles are regarded as responses to the news story. Thus, each news story in a topic thread structure should first be related to corresponding Web news articles.

For each pair of text from a Web news article and closed-caption (CC) from a news story, the following process is applied to determine if they should be related.

1. Apply morphological analysis[1] to each text.
2. Extract nouns and unknown words as keywords and make a term frequency vector for each text.
3. Calculate cosine similarity between both vectors.
4. If the cosine similarity exceeds a threshold θ_{rel}, the pair is determined as related.

[1]A Japanese morphological analyzer MeCab developed at Kyoto University and NTT was used. It is available for download from `http://taku910.github.io/mecab/`, Accessed on July 8, 2015.

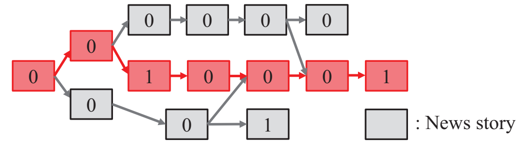


Figure 3: Selection of a topic thread (news story series) by maximizing the sum of binarized story scores.

2.3 Scoring terms in each topic cluster

Scores are assigned to terms based on the SNS responses to news stories in each topic cluster. In detail, terms that appear in each topic cluster are scored as follows.

1. Apply morphological analysis[1] to CCs of all news stories in a topic cluster.
2. Extract nouns and unknown words and make a TF-IDF vector for each news story.
3. The term score of term j in the i-th news story $W_{\mathrm{story}}(i,j)$ is defined as Eq. (1), where N is the number of SNS responses to a Web news article.

$$W_{\mathrm{story}}(i,j) = \begin{cases} W_{\mathrm{CC}}(i,j)N & (W_{\mathrm{article}}(i,j) \neq 0) \\ W_{\mathrm{CC}}(i,j) & (\text{otherwise}) \end{cases} \tag{1}$$

Here, $W_{\mathrm{CC}}(i,j)$ is the TF-IDF value of term j in the i-th news story, and $W_{\mathrm{article}}(i,j)$ is the frequency of term j in a Web news article related to the i-th story.

Also, the term score of term j in the k-th topic cluster $W_{\mathrm{cluster}}(k,j)$ is defined as

$$W_{\mathrm{cluster}}(k,j) = \textstyle\sum_{i \in \mathrm{C}_k} W_{\mathrm{story}}(i,j). \tag{2}$$

Here, C_k is a set of indices of news stories belonging to the k-th topic cluster.

2.4 Scoring sentences and news stories

Since we consider that news stories with many highly scored terms are important, news stories are scored based on the term scores. First, CC is divided to sentences. Next, sentence scores are calculated as the average of term scores in each sentence. Finally, the story score is calculated as the average of sentence scores in a news story.

2.5 Selection of a topic thread

According to story scores, each news story is assigned a binary value by thresholding (θ_{story}). Then a topic thread (news story series) which has the highest sum of the binary values is selected as the representative one. The binarization of story scores was performed in order to select a topic thread that includes a few highly scored news stories rather than that includes many averagely scored news stories. An example of this process is shown in Fig. 3.

(a) Indicator that shows the change of topic clusters.

(b) Indicator that shows the change of news stories by cropping the title of the story from the video.

Figure 4: Frames inserted to show the change of (a) topic clusters, and (b) news stories.

2.6 Generation of a video summary

We assume that objects or situations related to the terms that appear in the CC should also appear in the video. Thus, a video summary with a length of t [sec.] is generated according to sentence scores of all sentences in the selected topic thread, as follows.

1. Assign a video length t_k for the k-th topic cluster as
$$t_k = t \frac{|\mathrm{C}_k|}{\sum_i |\mathrm{C}_i|}. \tag{3}$$
2. In each topic cluster, reorder each sentence in descending order of sentence scores, and select sentences from the top so that the total video length corresponding to them should not exceed t_k [sec.]. Here, sentences with certain patterns that tend to appear in the beginning of a news story were discarded since corresponding video segments were usually composed of visually less interesting studio shots.
3. Reorder the sentences selected in each topic cluster according to the original time sequence.
4. Concatenate the video segments corresponding to the selected sentences. Here, a frame which indicates the topic cluster number is inserted at the change of topic clusters (Fig. 4 (a)) and the title of the news story with synthesized audio reading it out is inserted at the change of news stories (Fig. 4 (b)).

3. EXPERIMENT

3.1 Dataset

We prepared news stories obtained from a Japanese daily news show "NHK News7" broadcast during February and September 2014. Four topic thread structures constructed during this period (Data 1–4) were prepared. Table 1 shows the contents of the initial news stories, and Table 2 shows their statistics. We used "NHK NEWSWEB" [2] as a source of Web news articles, and the numbers of "like", "share", and "comment" of an SNS "Facebook" [3] posted to them as SNS responses. The SNS responses were collected one week after each news story was broadcast, which was also the timing that the news article on the Web site was removed.

[2] `http://www3.nhk.or.jp/news/`, Accessed on July 8, 2015.
[3] `https://www.facebook.com/`, Accessed on July 8, 2015.

Table 1: Contents of the initial news stories of the topic thread structures used in the experiment.

Data ID	Date	Contents
1	Feb. 21, 2014	Police and demonstration clashes in Ukraine
2	May 15, 2014	The right of collective defense discussed within the ruling party
3	May 22, 2014	Sanctions considered against North Korea's abduction
4	Sept. 12, 2014	Independence movement in Catalunya, Spain and Scotland, UK

Table 2: Statistics on the topic thread structures used in the experiment.

Data ID	1	2	3	4
Number of topic threads	67	383	223	4
Number of news stories	40	50	41	8
Number of topic clusters	15	15	11	1
Average number of Web news articles related to a news story	3.4	2.9	2.8	8.0

The threshold θ_{story} for the selection of a topic thread was set as the sum of the average and the standard deviation of all the story scores in a topic thread structure. The threshold θ_{rel} for relating a Web news article and a news story was set as 0.40.

3.2 Evaluation of news story series selection

First, the proposed topic thread or news story series selection method is evaluated.

3.2.1 Selection methods

In addition to the proposed method, two comparative methods for selecting a news story series were prepared as follows.

- Proposed method: News stories that compose a topic thread selected according to term scores using TF-IDF and SNS responses.
- Comparative method 1: News stories that compose a topic thread selected according to term scores using only TF-IDF.
- Comparative method 2: News stories in the topic thread structure whose similarity with the initial news story is no less than a threshold. This method is similar to the traditional topic clustering approaches.

3.2.2 Subjective evaluation

We conducted a subjective evaluation to confirm the effectiveness of the proposed news story series selection method. The subjects were fifteen male and two female students all in their twenties, among which ten subjects were assigned to evaluate each data. The subjects were asked to select the best news story series selected by one of the three methods,

Table 3: Subjective evaluation on the selected news story series.

<table>
<tr><th>Data ID</th><th>1</th><th>2</th><th>3</th><th>4</th><th>Avg.</th></tr>
<tr><td>Proposed method</td><td>4</td><td>9</td><td rowspan="2">8</td><td>6</td><td>6.75</td></tr>
<tr><td>Comparative method 1</td><td>4</td><td>1</td><td>4</td><td>4.25</td></tr>
<tr><td>Comparative method 2</td><td>2</td><td>0</td><td>2</td><td>0</td><td>1.00</td></tr>
</table>

from the point of the appropriateness of the overall flow of the news topic. Note that the news story series were presented in the same order to all the subjects, and the methods that they were selected by were not informed to the subjects.

3.2.3 Results and discussions

Table 3 shows the result; the number of subjects who selected the method as that selected the best news story series. Note that for Data 3, the result was combined since an identical news story series was selected by both the Proposed method and Comparative method 1.

As a result, the Proposed method obtained most votes for all four data. This owes to the effectiveness of the proposed topic thread selection method. Comparative method 2 obtained few votes because topics changed quite often. Meanwhile, the difference of the Proposed method and Comparative method 1 is the use of SNS responses. Since the Proposed method obtained more votes than Comparative method 1, we consider that considering the SNS responses was appropriate.

3.3 Evaluation of the generated video summary

Next, the proposed video summary generation method is evaluated.

3.3.1 Summarization methods

In addition to the proposed method, two comparative methods for generating the video summaries were prepared as follows.

- Proposed method: Summarization by concatenating video segments corresponding to each sentence selected according to term scores based on TF-IDF and SNS responses.
- Comparative method 1: Summarization by concatenating video segments corresponding to each sentence selected according to term scores based only on TF-IDF without SNS responses.
- Comparative method 2: Summarization by concatenating anchor shots that appear at the beginning of each story without using term scores. This method was prepared because usually the beginning of a news story called a lead-sentence introduces the summary of the news story. Although there are methods that could automatically detect anchor shots, here, we simply extracted them manually.

For the Proposed method and Comparative method 1, the maximum video length was set to $t = 180$ sec. For Comparative method 2, the maximum length was not determined, and simply all the corresponding anchor shots were concatenated.

Table 4: Lengths of videos generated by Comparative method 2.

Data ID	1	2	3	4
Length [sec.]	105	335	368	117

Table 5: Subjective evaluation on the generated video summaries.

Data ID	1	2	3	4	Avg.
Proposed method	3	**6**	3	**6**	**4.50**
Comparative method 1	**4**	4	**6**	1	3.75
Comparative method 2	3	0	1	3	1.75

3.3.2 Subjective evaluation

We conducted a subjective evaluation of the generated video summary with the same setup and in the same manner as in the previous experiment.

After watching the video summaries generated by different methods, the subjects were asked to select one that summarized the news story series the best.

Note that while the selection of sentences was performed by the proposed method automatically, the extraction of the title caption (Fig. 4 (a)), the insertion of frames (Fig. 4 (b)), and the detection of anchor shots were performed manually, although they could be implemented in the future by making use of existing techniques.

3.3.3 Results and discussions

Table 4 shows the lengths of the video summaries generated by Comparative method 2, and Fig. 5 shows the summary video generated by the proposed method for Data 4.

Table 5 shows the results from the subjective experiment. The Proposed method and Comparative method 1 which are based on sentence selection obtained more votes than Comparative method 2 that simply concatenates anchor shots. The subjects' comments were that the videos generated by Comparative methods 1 and 2 were "boring because there was no action", "redundant", or "not well explained because it was only composed of outlines". Next, the Proposed method obtained more votes than Comparative method 1 in average. The subjects' opinions were such as "a scene other than that of an anchorperson is better".

In fact, there were cases where an anchorperson and a commentator discussed the situation in a studio, so the persons, the places, or the situations which were mentioned in the CC did not appear, opposing our assumption. Although currently, the proposed method calculates sentence scores based on term scores, using image features in addition could improve the results. For example, detecting concepts (persons, places, objects, etc.) in video segments and combining them in the score calculation when selecting video segments to include in the summary could be effective.

4. CONCLUSION

In this paper, we proposed a method to generate a video summary on a news topic based on SNS responses to news stories. The proposed method selected a topic thread from a topic thread structure based on the term scores and gener-

Figure 5: Summary video generated by the proposed method for Data 4. In this case, there was only one topic cluster.

ated a summarized video by concatenating video segments corresponding to sentences with high scores in the selected topic thread.

Future work includes obtaining SNS responses in other sources. Handling SNS responses with different natures differently, taking into account linked documents and the contents in comments, and considering the timing of the SNS responses are other important tasks. We should also consider image contents when selecting video segments to be included in the summary, so that it should visually be informative as much as possible.

Acknowledgments

We thank the National Institute of Informatics, Japan for allowing us the usage of the news videos in the NII TV-RECS archive. Parts of this research were supported by MEXT, Grant-in-Aid for Scientific Research (23300036).

5. REFERENCES

[1] J. G. Fiscus and G. R. Doddington. Topic detection and tracking evaluation overview. In J. Allan (Ed.), *Topic Detection and Tracking: Event-based Information Organization*, Kluwer Academic Publishers, Dordrecht, pages 17–32. 2002.

[2] P. Duygulu, J. Pan, and D. A. Forsyth. Towards auto-documentary: Tracking the evolution of news stories. In *Proceedings of the 12th ACM International Multimedia Conference*, pages 820–827. ACM, October 2004.

[3] X. Wu, C.-W. Ngo, and Q. Li. Threading and autodocumenting news videos. *IEEE Signal Processing Mag.*, 23(2):59–68, March 2006.

[4] I. Ide, T. Kinoshita, T. Takahashi, H. Mo, N. Katayama, S. Satoh, and H. Murase. Efficient tracking of news topics based on chronological semantic structures in a large-scale news video archive. *IEICE Trans. on Information and Systems*, E95-D(5):1288–1300, May 2012.

[5] K. Kato, I. Ide, D. Deguchi, and H. Murase. Estimation of the representative story transition in a chronological semantic structure of news topics. In *Proceedings of the 4th ACM International Conference on Multimedia Retrieval*, pages 487–491. ACM, April 2014.

[6] E. Smits and A. Hanjalic. A system concept for socially enriched access to soccer video collections. *IEEE MultiMedia*, 17(4):26–35, October–December 2010.

[7] C.G.M. Snoek, B. Freiburg, J. Oomen, and R. Ordelman. Crowdsourcing rock n’ roll multimedia retrieval. In *Proceedings of the 18th ACM International Conference on Multimedia*, pages 1535–1538. ACM, October 2010.