

Multi-facet information processing algorithms for news video based on event combing

Shujin Lin¹, Fan Zhou², Guifeng Zheng², Kunhong Li², Mingjie Zhou²

¹School of Communication and Design

²School of Data and Computer Science, National Engineering Research Center of Digital Life
Sun Yat-sen University, Guangzhou, 510006, China

*Corresponding Author: linshjin@mail.sysu.edu.cn

Abstract—As an important and effective media of communication, online news videos are being widely and incessantly consumed by innumerable audience through the Internet. However, identifying content of interest from massive online resources remains a quite challenging and time-consuming task due to the continuous explosive growth of news video data, increasingly diverse production modes and the shortage of powerful tools for efficient video navigation, which are seriously affecting user's experience and information seeking efficiency. In this paper, a multi-facet information processing algorithm for news video based on event combing is reported. By establishing causal relationship and subordinate relationship between news events can help users comb news events and acquire the desired information from massive news video data both effectively and efficiently. The algorithms are mainly made up of three functional modules: the first one is the news timeline. The news videos is associated with their corresponding time nodes to reflect the times of occurrence of events. According to the sequence of news time nodes, users can better grasp the whole news events in the occurrence sequence of news events. The second is the news content keyword cloud module. It represents the content keywords of video within a specific time period, and helps users quickly appreciate the theme of video through the word cloud. The third is news topic clue clustering. According to the retrieval keywords entered by users, the system retrieves relevant videos from the database and carry out topic clustering operation to them. Every cluster is used to generate news clues in the form of card stack, reflecting the main content of the whole news event from different views. The advantages of the multi-facet information processing are demonstrated through benchmarked experimental results in comparison with peer methods.

Index Terms—News video, algorithm, content-based video clustering, multi-facet information processing

I. INTRODUCTION

News video has become a common way of news transmission. People can intuitively and vividly understand news events, hot topics and character dynamics through video. However, the existing video searching and browsing methods still have some limitations concerning concise and comprehensive organization and presentation of information [1–4]. The main problem is that the conventional video retrieval results are simply list the basic information such as thumbnails, titles and release date of the matched videos for a given query word, without further distinguishing the subtopics and time of occurrence of news events at a finer level of granularity. As a consequence, online news video viewers generally have to spend a great deal of time and effort identifying their content

of interest in the absence of informative content positioning clues and interactive search components.

Video structure analysis is one of the most important video processing techniques to facilitate the access of unstructured video content [5, 6]. Video files are hierarchically decomposed into multiple semantic subunits to further analyze video content and extract news keywords. But it is known that news video is different from other kinds of video. News has distinct timing, statement and relevance [7]. There are also important causal and subordinate relationships between news events. The functions of news video search and exploration can not only help users to quickly query and obtain news information, but also should help users to sort out and interpret the occurrence, continuity and function of news events, and grasp news information from a multidimensional perspective. For this purpose, the characteristics of news video itself can be considered to obtain more relevant information of news events in video news clustering, news content keywords and time node extraction. These are also concern issues researchers are paying attention to [8].

In this paper, a multi-facet information processing algorithms for news video based on event combing are designed to categorize news events to help users better understand the cause and effect of news events. Through event combining and clustering, topic clues and timeline of news events were established, which not only provided users with news video retrieval results, but also interpreted news events for users from multiple perspectives to more fully demonstrate the time of occurrence and relevance of events. The news timeline, news word cloud and clustering of news topic clues are provided to the user so that users can understand the news they are interested in more clearly. In the multi-facet information processing algorithms, the text information is first extracted from the video frame and audio of the news video using optical character recognition and speech recognition tool [9, 10], respectively. Combining the traditional LDA model [11] with the natural language characteristics of news reports, a new keyword extraction model is constructed to realize the content keyword extraction of news videos. Furthermore, topic clustering using the text obtained from aforementioned operations is carried out to provide users with textual summarization for content exploration. And news time nodes [12] and topic clues [13] are established. Integrated these processing results, users

can grasp the multi-dimensional information of news video and the relevance and timing of news video. Validated by carefully designed and conducted user studies, our news video system can systematically reflect the relevance and timing of news event, and can better satisfy users' understanding of news events.

The main contributions of this paper are as follows:

(1) Taking full advantage of the multi-channel feature of video, a method to calculate the visual information score of candidate keywords mainly based on audio channel information and supplemented by visual channel information is proposed;

(2) Considering the importance of video titles and the role of semantic features, this paper proposes a method to calculate the title relevance score of candidate keywords by calculating the semantic similarity between them;

(3) Emphasizing the importance of semantic features of video, a method for calculating the semantic importance score of candidate keywords based on graph theory is designed.

The rest of this paper is organized as follows. In Section II, we present a multi-facet information processing algorithms in detail. In Section III, we describe and analyze some experimental results. Section IV is conclusions.

II. THE MULTI-FACET INFORMATION PROCESSING ALGORITHMS

A. Overview of the Proposed Algorithm

The algorithm performs three main steps in its processing of news video. Initially, combining the LDA model with the natural language features of news reports, a new keyword extraction model is constructed to automatically analyze the main content of video news better and extract the content keywords of video news quickly and accurately. Secondly, the text clustering of video news content is carried out according to the news keywords. For the candidate set of news video keywords, the correlation degree among the words in it is calculated to generate news topics; then the news is divided into the topic clusters generated by the clustering according to the correlation degree of documents and words, so as to extract the topic clues of news. Thirdly, the time nodes of news events are used to generate a timeline, and then a time clue of news events is generated to show the relevance and timing of news. Each step will be described with detail in the following.

B. News video keyword extraction algorithm based on the improved LDA model

The advantage of the LDA model is the ability to mine deeper information. The goal is to find hidden document topics in the text and to add more general information between words and documents. But the method of extracting key words from LDA model also has some limitations. In the processing of topics, if the number of given topics is too large, the generated results of topics are not obvious. On the contrary, if the number of given topics is too small, some words may belong to multiple topics, affecting the accuracy of the results [11].

News video has its special characteristics [14]. News headlines have the function of revealing and clarifying news content. Moreover, in the document structure of news, the distribution position about key information of news events in the full text of news reports is also regular. Keywords in news reports belong to named entities in terms of word attributes, and the recognition of named entities can better help extract news video keywords. When LDA model is directly used to extract keywords, it is easy to ignore such important features as article structure and word distribution of news reports, thus affecting the accuracy of news content keywords. To solve this problem, we propose an improved LDA language model combining the characteristics of news reports to extract keywords from news content. The news headlines, news keyword locations, and named entities are considered in the model. The new model is a linear combination of four feature components. The coefficient weights of each feature component are obtained through training and learning to determine the final model.

The formula representation of the LDA model is:

$$\begin{aligned} Score_{lda_i} &= \sum_{n=1}^N P(w_i|z_n) P(z_n|d_m) \\ &= \sum_{n=1}^N P(w_i|z_n, \hat{\phi}_m) P(z_n|\hat{\theta}_m, d_m) \end{aligned} \quad (1)$$

where $\hat{\theta}_m$ and $\hat{\phi}_m$ are the posterior estimates of θ_m and ϕ_m obtained through Gibbs sampling in document d_m ; θ_m and ϕ_m are both Dirichlet distributions with corpus-level hyper-parameter vectors α and β correspondingly. The relevance between the news headline and word can be calculated in the following equation:

$$Score_{title_i} = \sum_{w_i \in title} 1 \quad (2)$$

The relevance about positional distribution of words can be represented as:

$$\begin{aligned} Score_{loc_i} &= \sum_{l=1}^t \left(\sum_{w_i \in s_l} 1 + \sum_{w_i \in s_{n-l}} 1 \right) \\ &+ \sum_{l=t+1}^{n-t-1} \sum_{w_i \in s_l} (1 - \lg l / \lg n) \end{aligned} \quad (3)$$

here t is the threshold as a crucial sentence. We set the value of t to three, which indicates that the first three sentences and the last three sentences of the news report are assumed to be decisive, and the words within these sentences should be scored highly.

Finally, if the word is a real-world object, then it can obtain an additional score value as below:

$$Score_{ner_i} = \sum_{w_i \in ner} 1 \quad (4)$$

The *ner* (named-entity recognition) set has been created in preprocessing.

Suppose that the features of each word in the document d_m are entirely independent of each other, and a news report can be represented by a set of words, that is, $d_m = \{w_1, w_2, \dots, w_n\}$. Then the aggregate score of each word $w_i (i \in \{1, 2, 3, \dots, n\})$ is the linear combination of its characteristic components, as shown in Equation (5).

$$S_{mi} = \omega_1 Score_{lda_{mi}} + \omega_2 Score_{title_{mi}} + \omega_3 Score_{loc_{mi}} + \omega_4 Score_{ner_{mi}} \quad (5)$$

Wherein, $\omega_1 + \omega_2 + \omega_3 + \omega_4 = 1$, and the first k words sorted by the score are chosen to be the keywords.

In order to determine the weight of each component, we implement the principal component analysis method [15]. First, we build an initial principal component model:

$$\begin{cases} F_1 = u_{11}v_1 + u_{21}v_2 + u_{31}v_3 + u_{41}v_4 \\ F_2 = u_{12}v_1 + u_{22}v_2 + u_{32}v_3 + u_{42}v_4 \\ \dots \\ F_m = u_{1m}v_1 + u_{2m}v_2 + u_{3m}v_3 + u_{4m}v_4 \end{cases} \quad (6)$$

where F_1, \dots, F_m are the $m (m \leq 4)$ correlation coefficients of the model, and $\{v_1, v_2, v_3, v_4\}$ is a tetrad (LDA, Title, Location, and *ner*). u_{ij} is the coefficient in the decision matrix, which is derived from the matrix composed of the eigenvectors of the correlation coefficient matrix. λ_i is the eigenvalue of the correlation coefficient matrix. The value of u_{ij} is correlated with the initial loading f_{ij} , the initial loading f_{ij} can be calculated by SPSS, and the mathematical expression of the coefficient of the decision matrix, u_{ij} , is:

$$u_{ij} = \frac{f_{ij}}{\sqrt{\lambda_i}}, i = 1, 2, \dots, m \quad (7)$$

On the basis of Equation (7), a comprehensive evaluation function of the principal component model is developed as follow:

$$F^{(v)} = \sum_{j=1}^m (\lambda_j / s) F_j = a_1 v_1 + a_2 v_2 + a_3 v_3 + a_4 v_4 \quad (8)$$

$$s = \lambda_1 + \lambda_2 + \dots + \lambda_m \quad (9)$$

where a_1, a_2, a_3 and a_4 represent the integrated significance of the four components v_1, v_2, v_3, v_4 in the principal component model accordingly. The mathematical expression can be written as:

$$a_i = \sum_{j=1}^m (\lambda_j u_{ij} / s) \quad (10)$$

Finally, the weights of the each component is obtained by:

$$\omega_i = a_i / \sum_{j=1}^4 a_j \quad (11)$$

C. Establishment time nodes for news events

The event nodes of news events reflect the event clues of news reports from different angles. Our algorithm uses the existing news knowledge base to extract the time nodes of news events. First, based the methods mentioned above, we extract the keyword set of the target news report, $Q_i = \{w_1, w_2, \dots, w_n\}$. Then we use the keyword set as the retrieval input to obtain related information of the news report from the news knowledge-base interface of the search engine, like "http://Google.search_news(query=Q, n, start, country_code)". Next, using regular expression, we obtain the news titles, $X = \{x_1, x_2, \dots, x_U\}$, and news time current sequences, $T = \{t_1, t_2, \dots, t_U\}$, of the first u news documents. Finally, from the exact time current sequences, the time nodes sequences can be gleaned. And the most frequent time node is deemed to be the time of the news incident.

D. News topic clue extraction based on clustering

News topic thread extraction is a clustering process and categorized as unsupervised learning. The result of clustering is a number of news topic clusters. $C_0 = \{k_1, k_2, \dots, k_n\}$ is defined as a set that contains all the keywords of the document set. $f(k_i)$ indicates that the number of times the word k_i appears in the entire document set, and $f(k_i \cap k_j)$ represents the number of times k_i and k_j co-occur in the document set. The formula for calculating the correlation between the k_i and k_j is as shown in Equation (12).

$$r_{k_i, k_j} = \frac{f(k_i \cap k_j)}{\max(f(k_i), f(k_j))} \quad (12)$$

$C_q = \{k_{q1}, k_{q2}, \dots, k_{qm}\}$ is defined as a topic cluster. The formula for calculating the correlation between the k_i and C_q is as shown in Equation (13).

$$r_{k_i, C_q} = \frac{r_{k_i, k_{q1}} + r_{k_i, k_{q2}} + \dots + r_{k_i, k_{qm}}}{m} \quad (13)$$

Based on the previous definition, we can design the algorithm as follows: (a) If there is no unfulfilled topic cluster currently and $C_0 \neq \emptyset$, then we choose the k_i in C_0 with the max r_{k_i, C_0} value to initialize a new target cluster with; (b) If $C_0 \neq \emptyset$ and the max r_{k_i, C_q} value is greater then the predefined threshold Q , then remove the k_i with biggest r_{k_i, C_q} value from C_0 to the target cluster C_q repetitively; (c) Repeat steps (a) and (b) until C_0 is empty; (d) Filter out all the clusters whose elements is less than three. Eventually, the topic thread set, $S = \{C_1, C_2, \dots, C_p\}$, has been accomplished.

After getting the topic clusters, we need to divide each news document into the corresponding topic cluster. Every element in set S , a keyword cluster, can be expressed as $C_q = \{k_{q1}, k_{q2}, \dots, k_{qm}\}$, and each k_{qi} in C_q indicates that the keyword k_i of C_0 exists in C_q or not. The news document can be expressed in the same vector space of C_q as $D_m = \{w_{m1}, w_{m2}, \dots, w_{mn}\}$; w_{mj} is calculated with TF-IDF algorithm. Since C_q and D_m are in the identical vector space, the correlation of them can be revealed by their cosine value,

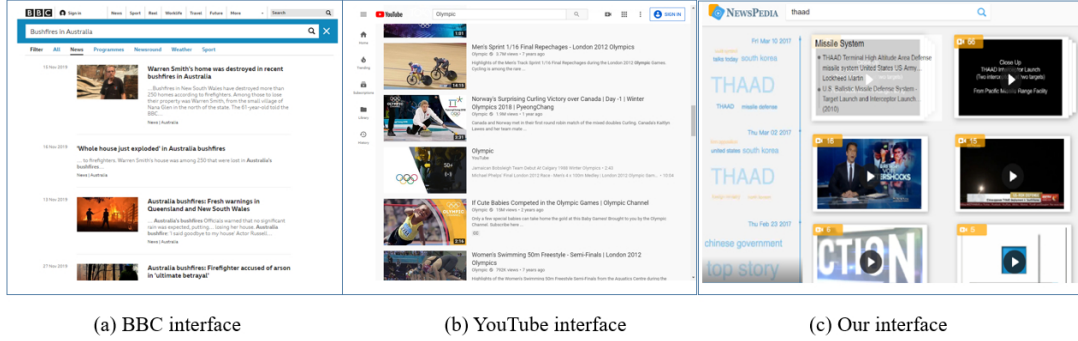


Fig. 1. Main interfaces about three systems

as is shown in Equation (14).

$$\text{Cor}(C_q, D_m) = \frac{\sum_{t=1}^n w_{mt} k_{qt}}{\sqrt{\sum_{t=1}^n w_{mt}^2 \times \sum_{t=1}^n k_{qt}^2}} \quad (14)$$

According to Equation (14), every news document can be assigned to a most relevant topic cluster. Together with the time nodes, the topic threads are obtained. The keywords in the cluster should be pertinent the tag of the topic thread.

III. USER STUDY

In order to test the efficiency of our algorithm in news video retrieval, we design an interface to display the results (timeline, topic event clusters, and word cloud).

We conducted a user study by comparing it with other two systems: BBC and YouTube. Among them, BBC is the largest news broadcaster in Britain and one of the largest in the world. YouTube system is one of the mainstream comprehensive video retrieval systems, with a large number of online users, and its news section can be used for news video retrieval. We chose the two modules that were most relevant to our work for a comparative study. Fig. 1 shows the main interfaces about this two referred system.

This paper refers to the system performance evaluation method proposed by Crabb [16] and Yadav [17]. These methods are mainly in the form of information search activities, which are commonly used in the literature to evaluate the effectiveness and efficiency of educational video exploration tools and systems. Two methods about objective evaluation based on task design and subjective evaluation based on user rating are designed to evaluate our system through comparative analysis of experiments. First, two video retrieval systems are selected for comparison with the system in this paper. In order to ensure the fairness and representativeness of the research experiment, 16 users were invited to participate in the system evaluation, whose ages are between 19 and 40 years old. All the surveyed users had experience of browsing English news videos on news video retrieval websites, so as to ensure a comprehensive performance evaluation of the system.

In generally, the goal of news video retrieval system is to help users quickly and accurately find what they want to know

about video news, and make the combining to the news event. Some relevant questions are set in the experiment. Users in the trials were asked to use three different video retrieval systems to find the answer for each question, and give out the experience of system evaluation, through the statistics and analysis of experiment responses from the user to evaluate system. The questions set up in the experiment can be divided into three categories:

- (1) News retrieval: The main purposes are to investigate the system can quickly and accurately find the target news video. (*Q1: Who was the main character in the email scandal? Q2: When does the closing ceremony of the Rio Olympics take place? Q3: What will former minister David Cameron focus on in his speech after the announcement of Britain's formal exit from the European Union?*)
- (2) News comb: It is used to investigate the effect of the system on news analysis. (*Q1: What is the sequence of events during the 2016 us election? Q2: What didn't happen at the 2017 Academy Awards? Q3: Which of the following news events are related to the Thaad incident?*)
- (3) User experience: These questions are used to the user's subjective judgment in terms of the overall satisfaction of the proposed system, the user-friendliness of the interface, and the reasonability of interactive design, etc. (*Q1: Do you think the functions of the system can help you complete the questions? Q2: Do you think the system is easy to understand and operate? Q3: Are you satisfied with the user interface design of the system? Q4: How satisfied are you with the system as a whole?*)

Fig. 2 shows the statistical chart of the time taken by users to answer news search and news comb questions. It can be seen that the user in response to the news comb type problem is time consuming than the overall answer type news search problem by time. But NewsPedia system in helping users answer news search problem type or news comb type, has the certain superiority. Overall, system NewsPedia takes the shortest time, and BBC system takes longest. The same in this figure, the average score of each system for all user experience questions. In terms of practicality, the system NewsPedia has highest score, which demonstrates its capability to help users search and comb video news. In terms of usability, the interfaces of

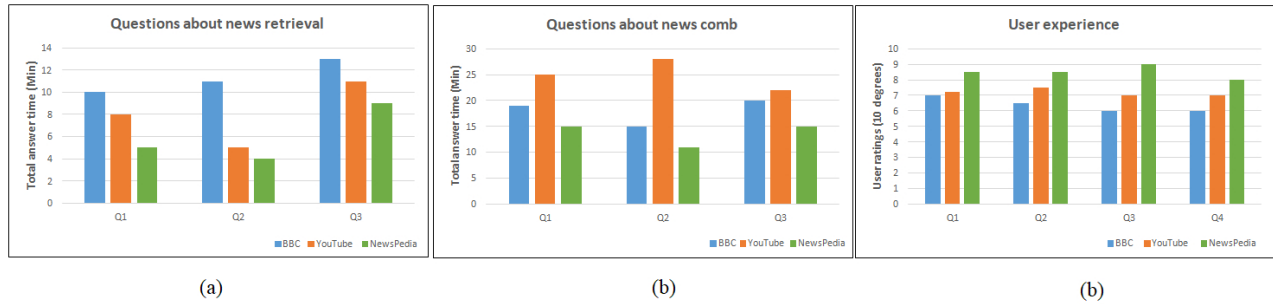


Fig. 2. User study by comparing with other two systems

NewsPedia and YouTube were rated higher than BBC. And these two systems have more functions, so they got higher scores on question 2 and 3. In terms of overall satisfaction, system NewsPedia has the highest score, since it helps users quickly find the target news videos accurately, and comb and understand the video news content, thus obtains the better user evaluation performance.

IV. CONCLUSION

The paper introduced a multi-facet information processing algorithm for news video based on event combing. The information about timeline, topic event clusters, and word cloud are integrated to help users acquire the news information efficiently from massive news video resources and comb news events. The experimental results suggest that the proposed system is efficient and effective in aiding users identify content of interest and understand news events.

ACKNOWLEDGMENT

The work is supported by the research and development plan in key areas of Guangdong Province (No.2020B010165001).

REFERENCES

- [1] Zhuo Lei, Chao Zhang, Qian Zhang, and Guoping Qiu, "Framerank: A text processing approach to video summarization," in *2019 IEEE ICME*, 2019, pp. 368–373.
- [2] Baoquan Zhao, Songhua Xu, Shujin Lin, Ruomei Wang, and Xiaonan Luo, "A new visual interface for searching and navigating slide-based lecture videos," in *2019 IEEE ICME*, 2019, pp. 928–933.
- [3] Da Cao, Zhiwang Yu, and Hanling Zhang et al., "Video-based cross-modal recipe retrieval," *ACM Multimedia 2019*, pp. 1685–1693, 2019.
- [4] Jakub Lokoc and Gregor Kovalcik et al., "A framework for effective known-item search in video," *ACM Multimedia 2019*, pp. 1777–1785, 2019.
- [5] Daniel Rotman, Dror Porat, and Gal Ashour, "Robust video scene detection using multimodal fusion of optimally grouped features," in *2017 IEEE 19th International Workshop on MMSP*, 2017, pp. 1–6.
- [6] Hongjie Chen, Lei Xie, Wei Feng, Lilei Zheng, and Yan-ning Zhang, "Topic segmentation on spoken documents using self-validated acoustic cuts," *Soft Comput.*, vol. 19, no. 1, pp. 47–59, Jan. 2015.
- [7] Dilek and Adnan, "Exploiting information extraction techniques for automatic semantic video indexing with an application to turkish news videos," *Knowledge-Based Systems*, vol. 24, no. 6, pp. 11–13, 2011.
- [8] Baraldi L and Grana C et al., "Recognizing and presenting the storytelling video structure with deep multimodal networks," *IEEE Trans. on Multimedia*, pp. 1–1, 2016.
- [9] Shivakumar K M, Aravind K G, and Anoop T V et al., "Kannada speech to text conversion using cmu sphinx," *ICICT*, pp. 19–20, 2017.
- [10] Yutong Gao, Feifan Song, Xiaqing Xie, Qing Sun, and Xu Wu, "Study of test classification algorithm based on domain knowledge," in *CCT*, Oct 2015, pp. 1–5.
- [11] Miha Pavlinek and Vili Podgorelec, "Text classification method based on self-training and lda topic models," *Expert Systems With Applications*, pp. 16–20, 2017.
- [12] Xujian Zhao, Peiquan Jin, and Lihua Yue, "Discovering topic time from web news," *Information Processing and Management*, vol. 51, no. 6, pp. 35–36, 2015.
- [13] Zhaoxin Fan, Shuoying Chen, Li Zha, and Jiadong Yang, "A text clustering approach of chinese news based on neural network language model," *Int. J. Parallel Program.*, vol. 44, no. 1, pp. 198–206, Feb. 2016.
- [14] S Foroozan, MA Azmi Murad, NM Sharef, and AR Abdul Latiff, "Improving sentiment classification accuracy of financial news using n-gram approach and feature weighting methods," in *ICISS. IEEE*, 2015, pp. 1–4.
- [15] Qifa Qu, "Determination of weights for the ultimate cross efficiency: A use of principal component analysis technique," *JSW*, vol. 7, no. 10, pp. 2177–2181, 2012.
- [16] Crabb M, Jones R, and Armstrong M et al., "Online news videos:the ux of subtitle position," *International ACM Sigaccess Conference*, pp. 215–222, 2016.
- [17] Yadav K, Gandhi A, and Biswas A et al., "Vizig:an anchor points based navigation and summarization in educational videos," *International Conference on IUI*, pp. 407–418, 2016.