

HIERARCHICAL NEWS VIDEO CATEGORIZATION BASED ON SEMANTIC ANALYSIS

Jia-Ching Wang, Chen-Yu Chen, Jhing-Fa Wang, and Chang-Ping Chen

Department of Electrical Engineering, National Cheng Kung University
1 University Road, Tainan, Taiwan, R.O.C.
E-mail: wangjf@csie.ncku.edu.tw

ABSTRACT

This study presents a hierarchical news video categorization architecture based on semantic analysis. First, news video is segmented into different news stories based on anchorperson shot detection. For each news story, we extract its caption information to perform semantic analysis. These stories are then semantically categorized as an adaptive hierarchy tree. This tree adaptive split its tree node in accordance with a term-based entropy measure. The superiority of the proposed system has been demonstrated through the three days news video obtained from EraNews in Taiwan.

1. INTRODUCTION

Due to the ever-changing society we live in, watching news has become a part of daily life to us. Always grasping news has already become a must for modern people to follow up the pace of the times. However, to most of people, they often do not have enough time and patience to finish watching long news. An effective news video browsing system based on news topic categorization is thus quite useful to them. In order to access the news selectively, the first step is to segment the news into different stories [1], [2]. As a result of the news story beginning with the anchorperson reporting, the news is segmented mostly by using anchorperson shots as segmentation boundaries. Therefore, extensive research has been published on anchorperson shot detection [3]-[5]. Since the news video can be segmented into stories, we thus can categorize these stories for improving the browsing convenience. Some of the early published works on news story categorization only can be seen the clue of adopting image feature for categorization [6], [7]. Because of the lack of using information, they could only categorize the news in a simplest structure. In later published research of news story categorization, the

caption information has become a basis of categorization [8]-[11]. Although these methods have achieved a certain measure of success in news story categorization, they still have some drawbacks. First, they only provide a flat structure, i.e. their categorization includes only one level. When the number of news stories related to a categorization item becomes grows, they don't have hierarchical structure to provide further categorization. Second, the items of their categorization are fixed. The fixed categorization causes a lack of flexibility regarding the various news events. To overcome the two drawbacks, a hierarchical tree structure based on unsupervised clustering is essential. Therefore, this paper develops an adaptive hierarchy tree for categorizing the news stories. This strategy is based on semantic analysis of the caption information. Further categorization is adaptively provided according to the entropy information of the tree node.

2. SYSTEM OVERVIEW

Figure 1 presents the concept illustrations of the proposed hierarchical news video categorization based on semantic analysis. In the architecture, we introduce adaptive hierarchy categorization to make news story browsing efficiently. The block diagram of our method is described in Fig. 2. Firstly, the video stream is divided into shots. In order to separate the individual news stories, anchorperson shots are detected and used as the news story boundaries. The obtained news stories are categorized semantically. After extracting and recognizing the caption, each news story has its own text information. The obtained text information is then utilized to retrieve related web pages. Besides the texts extracted from the video caption, the texts in the retrieved web pages are also used to represent a news story. We then transform these texts into a feature space so that each story is represented as a

feature vector. The news categorization process is based on an adaptive hierarchy tree. Moreover, a term-based entropy is used to determine if a node in this tree is a terminal node.

3. STORY SEGMENTATION AND TEXT INFORMATION EXTRACTION

It's the same as general video programs; news video also comprises individual video shots. This study uses mutual information and the joint entropy between frames to detect video shot boundaries [12]. To separate the individual news stories, anchorperson shots are detected and serve as the news story boundaries. That is, a news story comprises a series of shots beginning from an anchorperson shot and ending just before the next anchorperson shot. This study adopts an anchorperson shot detection algorithm presented by Chen et al. [13]. This detection algorithm is based on a novel similarity measure which considers the spatio-temporal variation through consecutive frames. The procedures for categorizing these separate news stories are discussed below.

The caption in the news video undoubtedly provides abundant semantic cues and can help explain the news content. This study thus utilizes the caption for semantic based news video categorization. To extract the caption information in a video sequence, the caption can be detected by [14]. The caption block is located using a bottom-up approach. This approach starts from local binarization between the rising and falling points in the image, followed by searching, filtering, feature extraction, and modified region growing of connected components. Some candidate blocks are then extracted for assigning weight to locate the caption block. After segmenting out the characters of the text, commercial Chinese OCR software is applied to recognize the segmented characters of the text.

4. STORY VECTOR CONSTRUCTION

The processing of the recognized caption is discussed next. News caption usually only provides key messages and a news story generally has few captions. To make the categorization of news stories be more correct, news captions are also used to extract the related news web pages. Restated, a news message is represented by including extracted news caption and related news web pages. Regarding the texts in the caption and related news web pages, they should be broken into separate terms. The term is selected as a phrase rather than a word for its syntactic significance. A text story can be represented as a feature vector, and each element of this vector

can be associated with a term. To find out more important terms, this paper used the TFIDF function [15] to analyze the importance of each term. The TFIDF function is defined as:

$$TFIDF(t_k, n_j) = TF(t_k, n_j) \cdot IDF(t_k), \quad (1)$$

where, $TF(t_k, n_j)$ denotes the term frequency representing the frequency of term t_k in the story n_j , and $IDF(t_k)$ represents the inverse document frequency representing the weight term t_k in the total stories.

If a term exists in many stories, the inverse document frequency is low. Restated the inverse document frequency is highest in the case where a term involving in only one story. The inverse document frequency is defined by

$$IDF(t_k) = \log \frac{N_s}{DF(t_k)}, \quad (2)$$

where N_s denotes the total number of stories and $DF(t_k)$ is the document frequency representing the number of stories in which the term t_k exists.

The term space can be reduced according to the TFIDF function from N'_f to N_f by selecting a subset of N'_f terms with highest TFIDF values.

These N_f terms can be used to represent a news story by a story vector $f_j = (f_{j1}, f_{j2}, \dots, f_{jN_f})$, where f_{jk} , $k = 1, 2, \dots, N_f$, is the term weight associated with the term t_k and the news story n_j . Because the lengths of all the new stories are different, this study uses normalized term frequency as the term weight. The normalized term frequency for term t_k in the story n_j is defined by

$$NTF(t_k, n_j) = TF(t_k, n_j) \cdot \left[\sum_k^{N_f} TF^2(t_k, n_j) \right]^{-1}. \quad (3)$$

5. ADAPTIVE HIERARCHY TREE

This work hierarchically categorizes chaotic news events on the basis of its semantic meaning so that readers can conveniently browse the news. A hierarchy tree is constructed by successively splitting tree nodes. Each tree node is split into its k children to form a k -ary tree based on the K-means algorithm. For browsing convenience, this work grows the hierarchy tree adaptively. Users will feel chaotic if the semantic relation among the stories in a node is too weak. An entropy based criterion can be used to determine whether a tree node will split or not. Small entropy value of a tree node indicates high intra-cohesiveness. Otherwise, the stories in this tree node are not tightly related. For a tree node TN, the stories

in TN are denoted as $S(TN) = \{s_1, s_2, \dots, s_{N_m}\}$, where N_m represents the story number of TN . The term-based entropy of TN is defined as

$$E_m = -\sum_{k=1}^{N_f} e_k \log e_k, \quad (4)$$

where $e_k = N_k/N_m$, and N_k denotes the number of stories in TN in which t_k occurs.

Each tree node splits if the corresponding term-based entropy is greater than a threshold. All the stories in the same tree node thus can be ensured to be closely related.

Table 1: Statistics for the tested news video program

Video clip name	Frame Rate	Total shot number	Total story number	Average shot number per story
Three-day EraNews	30 fps	698	45	15.5

Table 2: Evaluation of the news video categorization

Term-based entropy threshold	Weighted similarity	FScore measure
5	0.849	0.686
10	0.627	0.589
15	0.287	0.351
20	0.199	0.224

6. EXPERIMENTAL RESULTS

The news information selected for testing was obtained from EraNews Station in Taiwan. One hour of news per day was recorded over a three-day period. All of the video clips have 352×240 resolution at 30 fps. Table 1 lists the statistics for the three-hour news sampled.

The present structure level summarization is based on unsupervised clustering. To assess its effectiveness, the weighted similarity and the FScore measure are adopted [16], [17]. The cohesiveness of clusters can be used as a measure of cluster similarity. One method for computing the cluster cohesiveness is to adopt the weighted similarity of the internal cluster similarity. The FScore measure is another type of measures which lets us evaluate how well the clustering is working by comparing the groups produced by the clustering to known classes. Table 2 lists the experimental results. It is reasonable that both the weighted similarity and the FScore measure increases if we have a smaller term-based

entropy threshold. However, there is a tradeoff. Because a series of news stories related to the same topic will also be split into different clusters with a too small term-based entropy threshold.

7. CONCLUSION

With the demand of news video browsing and indexing, this study presents a semantic-based news video categorization method. The news video is segmented into different stories in accordance with the anchorperson shots. After extracting the caption information, each news story can be represented as a feature vector. These stories are then adaptively and hierarchically categorized based on semantic analysis. Future research directions for improving the proposed method include integration of audio cues and deeper natural language study.

REFERENCES

- [1] L. Chaisorn, T. S. Chua, and C. H. Lee; "The segmentation of news video into story units," in *Proceedings of IEEE International Conference on Multimedia and Expo*, 2002, pp. 73-76.
- [2] N. E. O'Connor, S. Marlow, N. Murphy, A. F. Smeaton, P. Browne, S. Deasy, H. Lee, and K. McDonald, "Fischlar: an on-line system for indexing and browsing broadcast television content," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 7-11 May 2001, pp. 1633-1636.
- [3] H. T. Chen, D. Y. Chen, and S. Y. Lee, "Object based video similarity retrieval and its application to detecting anchorperson shots in news video," in *Proceedings of Fifth International Symposium on Multimedia Software Engineering*, 2003, pp. 172-179.
- [4] X. Gao and X. Tang, "Unsupervised video-shot segmentation and model-free anchorperson detection for news video story parsing," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 9, pp. 765-776, Sep. 2002.
- [5] A. Hanjalic, R. L. Lagensijk, and J. Biemond, "Template-based detection of anchorperson shots in news programs," in *Proceedings of 1998 International Conference on Image Processing*, 4-7 Oct. 1998, pp. 148-152.
- [6] H. J. Zhang; Y. Gong, S. W. Smoliar, and S. Y. Tan, "Automatic parsing of news video," in *Proceedings of the International Conference on Multimedia Computing and Systems*, 15-19 May 1994, pp. 45-54.
- [7] S. Eickeler and S. Muller, "Content-based video indexing of TV broadcast news using hidden Markov models," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 15-19 March 1999, pp. 2997-3000.
- [8] W. Qi, L. Gu, H. Jiang, X. R. Chen, and H. J. Zhang, "Integrating visual, audio and text analysis for news video," in *Proceedings of 2000 International*

Conference on Image Processing, 10-13 Sept. 2000, pp. 520-523.

- [9] S. Raaijmakers, J. den Hartog, and J. Baan, "Multimodal topic segmentation and classification of news video," in *Proceedings of 2002 IEEE International Conference on Multimedia and Expo*, 26-29 Aug. 2002, pp. 33-36.
- [10] R. S. Jasinschi, N. Dimitrova, T. McGee, L. Agnihotri, J. Zimmerman, and D. Li, "Integrated multimedia processing for topic segmentation and classification," in *Proceedings of 2001 International Conference on Image Processing*, 7-10 Oct. 2001, pp. 366-369.
- [11] W. Zhu, C. Toklu, and S. P. Liou, "Automatic news video segmentation and categorization based on closed-captioned text," in *Proceedings of IEEE International Conference on Multimedia and Expo*, 22-25 Aug. 2001, pp. 829-832.
- [12] Z. Cernekova, C. Nikou, and I. Pitas, "Shot detection in video sequences using entropy based metrics," in *Proceedings of International Conference on Image Processing*, 24-28 June 2002, pp. III-421-III-424.

[13] H. T. Chen, D. Y. Chen, and S. Y. Lee, "Object based video similarity retrieval and its application to detecting anchorperson shots in news video," in *Proceedings of Fifth International Symposium on Multimedia Software Engineering*, 2003, pp. 172-179.

- [14] Y. K. Chen and J. F. Wang, "Locating the destination address block on images of complex mail pieces," *Journal of the Chinese Institute of Engineers*, vol.24, no. 6, pp. 761-770, Nov. 2001.
- [15] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information Processing and Management*, vol. 24, no. 5, pp. 513-523, 1988.
- [16] M. Steinbach, G. Karypis, and V. Kumar, "A comparison of document clustering techniques," in *Technical Report #00-034*, Department of Computer Science and Engineering, University of Minnesota.
- [17] Y. Zhao and G. Karypis, "Evaluation of hierarchical clustering algorithms for document datasets," in *Proceedings of the eleventh international conference on Information and knowledge management*, 2002, pp. 515-524.

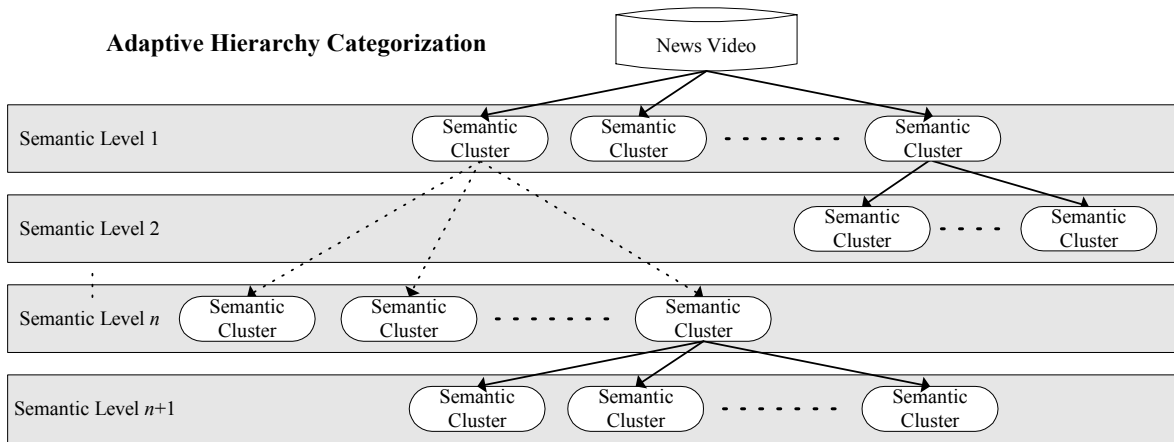


Fig. 1 Concept illustrations of the semantic-based adaptive hierarchy categorization for news video.

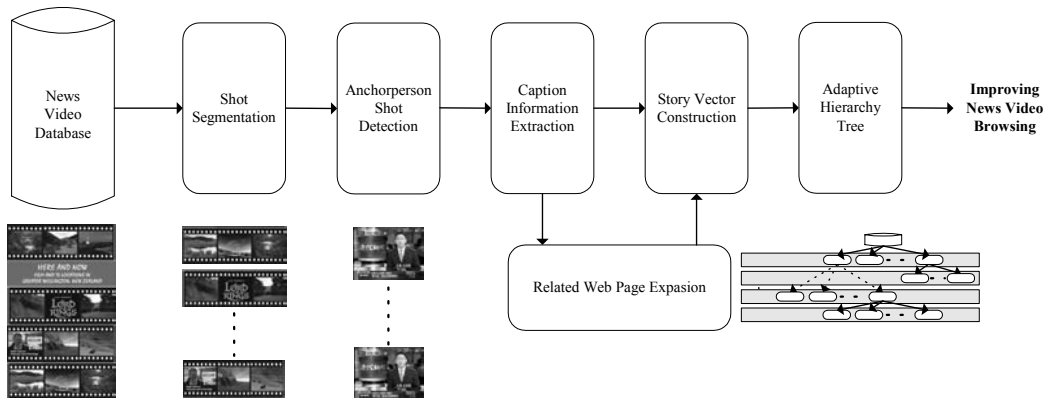


Fig. 2 Block diagram of the proposed news video categorization method.