

Summarization of Documentaries

Kezban Demirtas¹ Ilyas Cicekli² Nihan Kesim Cicekli¹

¹ Department of Computer Engineering, Middle East Technical University, Ankara, Turkey

² Department of Computer Engineering, Bilkent University, Ankara, Turkey
kezbandemirtas@gmail.com, ilyas@cs.bilkent.edu.tr, nihan@ceng.metu.edu.tr

Abstract. Video summarization algorithms present condensed versions of a full length video by identifying the most significant parts of the video. In this paper, we propose an automatic video summarization method using the subtitles of videos and text summarization techniques. We identify significant sentences in the subtitles of a video by using text summarization techniques and then we compose a video summary by finding the video parts corresponding to these summary sentences.

Keywords: Video Summarization, Text Summarization.

1 Introduction

Video content is being used in a wide number of domains ranging from commerce, security, education and entertainment. People want to search and find the video content according to its semantics. Creating searchable video archives becomes an important requirement for different domains as a result of the increase in the amount of multimedia contents. Video summarization helps people to decide whether they really want to watch a video or not. Video summarization algorithms present a condensed version of a full length video by identifying the most significant parts of the video.

In this paper, we propose an automatic video summarization system in order to present summaries to the users so that they can decide easily whether the selected video is of any interest to them. We aim to use text information only to determine how only the text data associated with the video is helpful in searching the semantic content of videos. The subtitles provide the speech content with the time information which is used to retrieve the relevant video pieces. For this purpose, we have chosen documentary videos as the application domain. In documentary videos, the speech usually consists of a monolog and it mentions the things seen on the screen.

For automatic summarization, we make use of two text summarization algorithms [1,3] and combine the results of these two algorithms to constitute a summary. Text summarization techniques identify the significant parts of a text to constitute a summary. We extract a summary of video subtitles with these summarization algorithms and then we find the video parts corresponding to these summary parts. By combining the video parts, we create a moving-image summary of the original video. In our summarization approach, we take the advantage of the documentary video characteristics. For example, in a documentary about “animals”, when an animal is seen on the screen, the speaker usually mentions that animal. So, when we find the video parts corresponding to the summary sentences of a video, those video parts are

closely related with the summary sentences. Hence we obtain a semantic video summary giving the important parts of a video.

Text features associated with a video can be viewable text placed on the screen or transcript of the dialog which can be provided in the form of closed captions, open captions or subtitles. Text features plays an important role in video summarization as it contains detailed information about the video content. Pickering et al. [4] make summarization of television news by using the accompanying subtitles. They extract news stories from the video and provide a summary for each story by using lexical chain analysis. Tsoneva et al. [5] creates automatic summaries for narrative videos using textual cues available in subtitles and scripts. They extract features like keywords, main characters' names and presence, and according to these features they identify the most relevant moments of video for preserving the story line. In our video summarization system, we extract moving-image summaries of documentaries using video subtitles and text summarization methods.

The rest of the paper is organized as follows. Section 2 describes our video summarization approaches and we present evaluations of these approaches in Section 3. Finally in Section 4, conclusions and possible future work are discussed.

2 Video Summarization

We find the summary sentences of the subtitle file by using the text summarization techniques [1,3]. Then we find the video segments corresponding to these summary sentences. By combining the video segments of summary sentences, we create a video summary. Subtitle files contain the text of the speech, the number and time of speech. In the text preprocessing step, the text in the subtitle file is extracted by striping the number and time of the speech, and it is given to the "Text Summarization" module. "Text Summarization" module finds the summary sentences of the given text. We use three algorithms for finding the summary sentences; TextRank algorithm [3], Lexical Chain algorithm [1] and a combination of these two algorithms. After the summary sentences are found by one of these approaches, the output can be given to the "Text Smoothing" module. This module applies some techniques to make summary sentences more understandable and smoother. "Video Summarization" module creates the video summary by using the summary sentences. This module finds the start and end times of sentences from the video subtitle file. Then the video segments corresponding to start and end times are extracted. By combining the extracted video segments, a video summary is generated.

The TextRank algorithm [3] extracts sentences for automatic summarization by identifying sentences that are more representative for the given text. To apply TextRank, we first build a graph and a vertex is added to the graph for each sentence in the text. To determine the connection between vertices, we define a "similarity" relation between them, where "similarity" is measured as a function of their content overlap. The content overlap of two sentences is computed by the number of common tokens between them. To avoid promoting long sentences, the content overlap is divided by the length of each sentence.

In [1] automated text summarization is done by identifying the significant sentences of text. The lexical cohesion structure of the text is exploited to determine the importance of sentences. Lexical chains can be used to analyze the lexical

cohesion structure in the text. In the proposed algorithm, first, the lexical chains in the text are constructed. Then topics are roughly detected from lexical chains and the text is segmented with respect to the topics. It is assumed that the first sentence of a segment is a general description of the topic, so the first sentence of the segment is selected as the summary sentence.

We also propose a new summarization approach by combining the two summarization algorithms, TextRank algorithm [3] and Lexical Chain algorithm [1]. In this approach, we find the summary sentences of a text by using both the TextRank algorithm and the Lexical Chain algorithm. Afterwards, we determine the common sentences of two summaries and select these sentences to be included in the summary. Both algorithms determine the summary sentences of a text in a sorted manner, that is, the summary sentences are sorted with respect to their importance scores. After selecting the common sentences, we select the most important sentences of the two algorithms up to the length of the desired summary.

In order to improve the understandability and completeness of the summary, some smoothing operations are done after text summarization. It is observed that some of the selected sentences start with a pronoun and if we do not have the previous sentences in the summary, these pronouns may be confusing. In order to handle this problem, if a sentence starts with a pronoun, the preceding sentence is also included in the summary. If the preceding sentence also starts with a pronoun, its preceding sentence is also added to the summary sentence list. The backward processing of the sentences goes at most two steps. We observed that if a sentence starts with a pronoun, including just the preceding sentence solves the problem in most cases and the summary becomes more understandable.

3 Experiments and Evaluation

The evaluation of video summaries is a hard job because summaries are subjective. Different people will compose different summaries for the same video. The evaluation of video summaries could be conducted by requesting people watch the summary and asking them several questions about the video. However, in our summarization system, since we use text summarization algorithms, we prefer to evaluate the text summarization algorithms only. We believe that the success of the text summarization directly determines the success of video summarization in our system. For the evaluation of text summarization, we use ROUGE (Recall-Oriented Understudy for Gisting Evaluation) algorithm [2] which makes evaluation by comparing the system generated output summaries to model summaries written by humans.

In our video summarization system, we tried six algorithms (three text summarization algorithms with or without smoothing the result) by using the documentaries from BBC. We asked students to compose summaries of the selected documentaries by selecting the most important twenty sentences from the subtitles. The same documentaries were also summarized by our video summarization system which generated summaries composed of twenty sentences by using our algorithms. In order to compare the system outputs with human summaries, the ROUGE scores are calculated, and given in [Table 1](#). From [Table 1](#), we can observe that smoothing improves the performance of all the algorithms. Our best method is the combination

of two algorithms using smoothing, and our best scores are comparable with the scores of the state of the art systems in the literature.

Table 1. ROUGE Scores of Algorithms in Video Summarization System

<i>Summarization Algorithm</i>	<i>ROUGE-1</i>	<i>ROUGE-L</i>	<i>ROUGE-W</i>
TextRank	0,33877	0,33608	0,13512
TextRank Smooth	0,34453	0,34184	0,13686
LexicalChain	0,24835	0,24600	0,10413
LexicalChain Smooth	0,25211	0,24976	0,10529
Mix	0,34375	0,34140	0,13934
Mix Smooth	0,34950	0,34716	0,14108

4 Conclusions

This paper presents a system which performs automatic summarization of documentary videos with subtitles. We perform video summarization by using video subtitles and employing text summarization methods. In this work, we take the advantage of the characteristics of the documentary videos. In documentary videos, the speech and the display of the video have a strong correlation in the way that mostly both of them give information about the same entities.

In the evaluation of video summaries, we evaluate the text summaries of videos. We compare the program summaries with human generated summaries and find the ROUGE score of program summaries. As a future work, we want to perform the detailed user evaluation of video summaries. Video summaries could be watched by viewers and the viewers could evaluate the results.

Acknowledgments

This work is partially supported by The Scientific and Technical Council of Turkey Grant ‘‘TUBITAK EEEAG-107E234, and The Scientific and Technical Council of Turkey Grant ‘‘TUBITAK EEEAG-107E151’’.

References

1. G. Ercan, and I. Cicekli. Lexical cohesion based topic modeling for summarization. In Proceedings of the CICLing 2008, pp. 582–592.
2. C.Y. Lin, and E.H. Hovy. Automatic evaluation of summaries using n-gram co-occurrence statistics. In Proceedings of HLT-NAACL-2003, Edmonton, Canada, 2003.
3. R. Mihalcea, and P. Tarau. TextRank - bringing order into texts. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2004), Barcelona, Spain.
4. M. Pickering, L. Wong, and S. Ruger. ANSES: Summarization of News Video. In Proceedings of CIVR-2003, University of Illinois, IL, USA, July 24-25, 2003.
5. T. Tsoneva, M. Barbieri, and H. Weda. Automated summarization of narrative video on a semantic level, In Proceedings of the International Conference on Semantic Computing, pp.169-176, September 17-19, 2007.