A Text-based Method for Detection and Filtering of Commercial Segments in Broadcast News

Ganesh Ramesh*, **Amit Bagga**†

*Department of Computer Science, University at Albany, SUNY,
1400, Washington Avenue, Albany, NY 12222, USA.
ganesh@cs.albany.edu

†Avaya Labs Research,
233 Mt. Airy Road, Basking Ridge, NJ 07920,USA.
bagga@avaya.com

Abstract

Story segmentation is an important problem in multimedia indexing and retrieval and includes detection of commercials as one of its component problems. Commercials appear regularly in television data and are usually treated as noise. Hence, filtering of commercials is an important task. This paper presents a system that detects and filters commercials from broadcast news data. While previous work in the area relies largely on features from audio, video and captions, the system described in this paper uses just closed caption text to perform this task. An evaluation of this system is also presented which shows comparable performance with other methods.

1. Introduction

The growing popularity of the Internet is allowing users instantaneous access to large amounts of information. Traditionally, this information has mostly been in text form. However, recent advances in telecommunications like DSL and cable modem services have greatly increased the bandwidth available to the common person, thereby allowing convenient access to much richer sources of information, which include audio and video. Therefore, the amount of multimedia data available over the Internet has increased significantly. In addition, traditional TV is also moving towards becoming digital and interactive. While multimedia data provides rich sources of information, its indexing and retrieval remains a significant challenge.

One of the main problems in multimedia indexing and retrieval is story segmentation which includes the detection of commercial segments that appear regularly in television data. Commercials are usually treated as noise and therefore need to be eliminated before indexing. This paper presents a system that detects and filters commercials from broadcast news data. While previous work in commercial detection used features from the video, audio, and closed caption text streams, the system described in this paper uses just the closed caption text. The paper also presents an evaluation of the system.

2. Related Research

Related research in commercial detection and filtering lies deeply embedded within the topic of story segmentation in broadcast news. (Hauptmann and Witbrock, 1998) deals with the goal to build a fully automatic system that could extract story boundaries using audio, video and closed-captioning cues. The commercial segments are detected using image features. Two features namely, the presence of black frames and the rate of scene changes are used in an ad hoc heuristic that performs the commercial detection. A hypothesis is made about which segments are commercial segments and they are revisited to make relabellings if needed. As the main focus of the paper was story segmentation, no quantifiable stand-alone performance measure was provided for commercial detection.

(Liu et al., 1998a) and (Liu et al., 1998b) deal with two approaches using neural networks and hidden markov models, respectively, to the problem of discriminating *five* types of TV programs, one type of which was commercials. Both these approaches use complex audio features to perform the task. Four types of classification results are studied in (Liu et al., 1998b) with a maximum classification performance of 74.5%. The performance of classification improved to 87.4% in (Liu et al., 1998a) using a 5−state HMM with 128 symbols. (Lu et al., 2001) studies an alternate approach which is based on a HMM but using chromaticity signatures from video summarization and their temporal relationship to perform the same discrimination task. However, they achieve a classification performance of only 66.7% for commercials.

(Wang et al., 2000) deals with content analysis and presents algorithms for segmentation and classification for video archiving and retrieval. A variety of features are used for classification, which include audio, color and motion features. Two different methods are used to combine these features for classification. The performance of commercial segment classification ranged from a low of 64.45% (for methods based solely on motions features) to a high of 93.58% (for the product approach that combined all the features together).

All the approaches discussed above use complex audio and video features to perform the task. While various forms of information are put to use, they tend to be time consuming and have an overhead in processing, often resulting in low scalability. The main motivation for our approach comes from trying to determine how well one can perform using **only** the data from closed captions and how useful this textual data can be in identifying commercials.

3. System Description

An analysis of the closed caption streams from a few broadcast news programs helped us make the following two observations:

1. Commercial segments in broadcast news are often preceded by cues that point to a commercial break. Examples are phrases like “coming up next”, “Still later tonight”, “When we come back,” and so on.

   We refer to each such cue phrase as a **key-phrase**. The collection of key-phrases is called a **key-phrase dictionary**. The key-phrases are fairly general and not specific to any particular channel.

   Therefore, the locations of the key-phrases within the closed caption stream identify the start of a potential commercial segment. Once the locations of the key-phrases are identified, the end of the potential commercial segment can be determined using the segment delimiter “>>” which denotes the start of a new segment or a speaker change.

2. Each commercial segment contains very disparate information when compared to the information content of the segments immediately preceding and following it. This is mainly because each commercial segment contains several, usually unrelated, commercials.

The observations listed above along with the text tiling approach in (Hearst, 1997) suggest the following approach. Consider the following three segments of closed caption text: a potential commercial segment, a segment immediately preceding it, and a segment immediately following it. Furthermore, partition each of the three segments into windows of equal temporal length. Finally, compute for each segment the average inter-window similarity between the closed caption text corresponding to the windows in that segment. Our observations lead us to believe that if the potential commercial segment does in fact contain commercials then its average inter-window similarity will be significantly lower than those of the surrounding segments.

However, there are two additional considerations that need to be described:

1. Elimination of non-commercial text that appears after key-phrases just before the start of an actual commercial segment.

   Almost all key-phrases in the key-phrase dictionary are followed by a very short announcement of the stories that will be covered later in the program. This announcement “advertises” upcoming stories and precedes the actual start of the commercials. Therefore, in order to determine the starting point of the commercials themselves, we need to eliminate this announcement. A time interval is used to filter out such instances of non-commercial text.

2. Elimination of multiple occurrences of **key-phrases** preceding a commercial segment.

   As described above, there is usually an announcement advertising upcoming stories prior to the actual start of commercials. Often, this announcement contains additional mentions of key-phrases that are part of the key-phrase dictionary. In such cases we use the key-phrase mentioned closest to the commercials.

Given: A document of **closed captions**, a **key-phrase dictionary**.

Output: A list of closed captions segments in the document that consist of commercials.

Steps:

1. Identify all the potential commercial segments using the key-phrases and the segment delimiter “>>”.
2. **For each** potential commercial segment
   - (a) Identify two equal segments of text that surround it.
   - (b) Split the potential commercial segment, and each of the surrounding segments in windows of equal temporal length.
   - (c) For each of the surrounding segments and the potential commercial segment, compute the average inter-window similarity of the windows within that segment. Section 3.2. describes how the similarity between two windows is computed.
   - (d) If the average inter-window similarity of the potential commercial segment is significantly lower (by a specified threshold) than either of the surrounding segments, then positively identify the potential commercial segment as a commercial segment.

   end for

Figure 1: **Algorithm Identify-Commercials**

3.1. Algorithm

The steps in figure 1 describe the algorithm used for identifying commercial segments, based on the approach described earlier.

3.2. VSM-Similarity

As mentioned earlier, each segment (either the potential commercial segment, or the two segments surrounding it) is divided into windows of equal temporal length. The vector space model (Salton, 1989) computes the average inter-window similarity for each segment. Therefore, for a given segment, the closed caption text associated with every window is input to this module. For the remainder of this section, we will refer to the closed captions associated with every window simply as “window.” The module compares every window to every other window in the segment.

In the vector space model, each window is stored as a vector of terms. The terms in the vector are in their morpho-

Station	# Programs	# Commercials (Truth)	Precision	Recall
ABC	10	29	28/29 = 97	28/29 = 97
CBS	16	56	47/52 = 90	47/56 = 84
NBC	13	39	37/41 = 90	37/39 = 95
Total	39	124	112/122 = 92	112/124 = 90

logical root form obtained using Porter's stemmer (Porter, 1980), and are filtered for stop-words. If S_1 and S_2 are the vectors for the two windows, then their similarity is computed as

$$Sim(S_1, S_2) = \sum_{common\ terms\ t_j} w_{1j} + w_{2j},$$

where t_j is a term present in both S_1 and S_2, w_{1j} is the weight of the term t_j in S_1 and w_{2j} is the weight of t_j in S_2.

The weight of a term t_j in the vector S_i for a window is given by

$$w_{ij} = \frac{\log \frac{N}{df} \times tf \times (k1 + 1)}{k_1 \times ((1 - b) + (b \times nsl(S_i))) + tf},$$

where tf is the frequency of the term t_j in the window, and N is the total number of non-empty windows in the segment being examined. df is is the number of windows in the segment that the term t_j occurs in. $nsl(S_i)$ is the average length of a window in the segment. k_1 and b are constants which are set to 2 and 0 respectively.

For each pair of windows, the vector space model returns a real number ≥ 0 that represents the similarity between the windows where 0 implies that there is nothing in common between them and progressively larger values imply greater similarity.

4. Experimental Results

We tested the algorithm on 39 broadcast news programs consisting of 124 commercial segments recorded from the ABC, CBS, and NBC stations in the USA. These programs were recorded over three weeks, one in February 2001, and two in October 2001. We used four other programs from the three stations for training the system, mainly for the creation of the key-phrase dictionary which contains 22 key-phrases. The bulk of these programs were 30 minutes long. However, two of them were 60 minutes long.

The length of each segment surrounding a potential commercial segment was set to 90 seconds. The window size for all the three segments was set to 15 seconds. Therefore, while the number of windows in each potential commercial segment varied according to its length, the number of windows in each surrounding segment was 6. The threshold factor for the inter-window similarity was set to 1.5. In other words, if the average inter-window similarity of either of the surrounding segments was 1.5 times the average inter-window similarity of a potential commercial segment, the algorithm would assume that the potential commercial segment actually contains commercials.

Figure 3.1. provides details about the performance of the system. The overall performance of the algorithm was 92% precision and 90% recall. This is just below the best performance achieved by (Wang et al., 2000) of 93.58%.

An analysis of the results showed that of the 12 missed commercial segments, 6 were missed because of the lack of the presence of a leading key-phrase while 6 were missed because the average inter-window similarity of either of the surrounding segments was not 1.5 times the average inter-window similarity of the commercial segment. In addition, the algorithm mis-classified 10 program segments as commercial segments. Almost all of these were mis-classified because of the nuances of using closed caption data (described in the next section), and the use of key-phrases to advertise either remaining stories in the current news program, or programs following the news program.

4.1. Limitations

One of the main limitation of our algorithm is that it relies heavily on the key-phrase dictionary for the detection of commercials. While the use of such key-phrases is prevalent in the USA, it may not be so in other countries. In addition, as evident in the analysis of the results, there are occasions when key-phrases are simply not used. In these cases, the system just cannot identify the commercial segments. Six out the 124 commercial segments in the data used for testing our system were not preceded by the use of a key-phrase. Moreover, the use of key-phrases is not restricted to commercial segments. They are also used to advertise programs following the news program and, at times, are used in the news program to advertise upcoming stories and are not followed by commercial segments. While the algorithm does a very good job of identifying a majority of these cases, it does mis-classify a few of these as commercial segments.

In addition to the design limitations of the algorithm, there are also limitations imposed by using only closed caption data. Closed caption data is far from perfect and, in general, consists of noise in the form of mis-spelled words, mis-recognized words, missing captions for some stories, and, finally, missing captions for commercials. In addition, speaker change and story boundaries (usually identified by $>>$ and $>>>$ signs) are not consistently used. A large number of the mis-classification errors are caused by the sparsity of the closed caption stream together with the presence of a key-phrase.

5. Conclusion

We have developed a simple yet novel algorithm that filters commercials from broadcast news data using only closed-caption text. The results obtained are comparable to those achieved using more sophisticated techniques and generalize reasonably to different channels.

The method proposed in this paper can be extended in several different ways to improve its general applicability and performance. In order to further improve the accuracy of the current method, it can be used as a pre-processing step for a procedure that factors in audio and video features. The pre-processing step will help reduce processing overhead. Finally, machine learning techniques can be used to classify the commercial segments.

6. References

Alexander G. Hauptmann and Michael J. Witbrock. 1998. Story segmentation and detection of commercials in broadcast news video. In *ADL-98 Advances in Digital Libraries Conference*, Santa Barbara, CA, April.

Marti A. Hearst. 1997. Texttiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics*, 23(1), March.

Z. Liu, J. Huang, and Y. Wang. 1998a. Classification of tv programs based on audio information using hidden markov model. Technical Report TR98-UT-LIPS-AGENTS-01.

Z. Liu, Y. Wang, and T. Chen. 1998b. Audio feature extraction and analysis for scene segmentation and classification. *Journal of VLSI Signal Processing System*, June.

Cheng Lu, Mark S. Drew, and James Au. 2001. Classification of summarized video by using hidden markov models on compressed chromaticity signatures. In *ANNIE'01: Artificial Neural Networks In Engineering*, St. Louis, Missouri, November.

M. F. Porter. 1980. An algorithm for suffix stripping. *Program*, 14(3).

G. Salton. 1989. *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Addison-Wesley.

Yao Wang, Zhu Liu, and Jin-Cheng Huang. 2000. Multimedia content analysis using both audio and visual cues. In *IEEE Signal Processing Magazine*, November.