Topic Segmentation in a Media Watch System

Rui Amaral[1,2,3] and Isabel Trancoso[1,3]

[1]Instituto Superior Técnico
[2]Instituto Politécnico de Setúbal
[3] L^2F - Spoken Language Systems Lab, INESC-ID
{Rui.Amaral,Isabel.Trancoso}@l2f.inesc-id.pt
https://www.l2f.inesc-id.pt

Abstract. This paper describes our on-going work on the topic segmentation module of a media watch system. The current version explores not only the typical structure of a broadcast news show, but also its contents, which are automatically produced by the speech recognition module, and the topic indexation module. The performance of the automatic topic segmentation module was compared with the manual segmentation done by a professional media watch company, yielding quite satisfactory results.

1 Introduction

Topic segmentation plays an important role in the prototype system for selective dissemination of Broadcast News (BN) in European Portuguese, developed at INESC-ID. The media watch system was initially built in the context of the ALERT European project [1] and is the object of continuous improvement in the framework of national project TECNOVOZ. The topic segmentation module (TS) described in this paper is one of the modules of the complex system and is performed off-line, exploring only audio-derived cues, for the time being. This paper starts with a brief description of our BN corpus in Section 2. The bulk of the paper is devoted to the topic segmentation module (section 3). Section 4 compares the automatic with the manual topic segmentation performed by a media watch company, and discusses the importance of video-derived cues. The final Section concludes and presents directions for future research.

2 The European Portuguese BN Corpus

The European Portuguese BN corpus includes different types of news shows, national and regional, generic and specific domains, from morning to late evening. In this work, we used 4 subsets, all manually segmented into stories, covering a wide range of scenarios. The SR (Speech Recognition) corpus contains 57h of BN shows, where 45% is presented by the lead anchor and the remaining shows also have a sports anchor. The JE (Joint Evaluation) corpus contains 13h, half of which contain only a lead anchor and the other half also include a sports anchor. To expand the segmentation scenarios, an extra BN corpus (EB) with 4h was

collected from a different TV station. One of the shows is presented by the lead anchor, but includes a local news commentator. The other two shows have two lead anchors, and one of them also includes the local news commentator. The need for the comparison with a professional media watch company motivated the collection of a very recent corpus (RTP07). This corpus contains around 6h, segmented by the media watch company. All the 6 shows have one lead anchor, without thematic anchors.

3 Topic segmentation

The goal of TS module is to split the BN show into its constituent stories, exploring their characteristic structure [2]. All stories start with a segment spoken by the anchor, and are typically further developed by out-of-studio reports and/or interviews. The analysis of the typical structure of a BN show led us to train a CART (Classification and Regression Tree) with potential characteristics for each segment boundary [3]. The CART performed reasonably well for BN shows with one lead anchor, but failed with shows involving 2 lead anchors. This led us to adopt a two-stage supervised approach: in a first stage of re-clustering, the two speaker ids with the most frequent turns are clustered into a single label. After this pre-processing stage, the CART is applied.

3.1 Exploring the topic related structure

To deal with a more complex structure, such as a BN show with a thematic anchor, a multi-stage approach was adopted where topic segmentation and indexation are interleaved. The first stage identifies potential story boundaries in every non-speech/anchor transitions. The second stage uses the topic indexation to isolate the thematic portion of the BN show (sports). This stage allows potential story boundaries to appear within the given theme. A third stage of boundary removal is applied using the same rules adopted by the CART. The knowledge of the topic was also used to remove false alarms in the weather forecast topic, which was typically split into multiple stories, due to the relatively long pauses made by the anchor between the forecasts for each part of the country.

3.2 Exploring non-news information

One recent improvement of our system is the inclusion of a non-news detector which detects the jingles that delimit the BN show, the publicity segments, and the headlines/teasers. The performance of the previous version of the TS module was seriously degraded by the presence of headlines [3], causing false alarms inside headlines, and miss boundaries after the headlines. The inclusion of the non-news information in the TS module allowed us to define another story boundary detection rule which avoided these problems.

3.3 Exploring the contents of BN segments

The main remaining problem was the false alarm rate due to the long anchor interventions in the middle or at the end of a story. In order to decrease these false alarms, we used the automatic transcriptions of the BN shows. The merging of short stories with either their left or right neighbors was dictated by a CART trained with the following features: the acoustic background conditions of the left and right stories, the word rate (computed at the first 7s of the short story, which is the minimal time required for a story introduction), the duration of the anchor segment, and the normalized count of matches of unigrams, bigrams and trigrams between the short story and the two neighbors. The matches are computed over the automatic transcripts and the purpose is to detect text similarities between the short story and its neighbors, to help the merge decision.

4 Results and discussion

Approach	%Recall	%Precision	F-measure	corpus
Single-Stage	79.6	69.8	0.74	JE
Two-Stage	81.2	91.6	0.85	EB
Multi-Stage	88.8	56.9	0.69	JE
Multi-Stage	97.1	86.8	0.92	RTP07-1
Multi-Stage (+meteo)	97.1	89.2	0.93	RTP07-1
Multi-Stage	98.9	71.7	0.83	RTP07-3
Multi-Stage + non-news info	96.8	73.9	0.84	RTP07-3
w/o ASR (eval=1s)	88.0	81.7	0.85	RTP07
with ASR (eval=1s)	91.2	83.0	0.87	RTP07
w/o ASR (eval=2s)	93.8	87.1	0.90	RTP07
with ASR (eval=2s)	97.0	88.2	0.92	RTP07

Table 1. Topic segmentation results.

The results of the different versions of the segmentation algorithm are presented in Table 1. The performance of the 3-stage approach only took the sports topic splitting into account (third line). The next two lines used the single BN show of RTP07 which had weather forecast news (RTP07-1). The fourth line taked only the sports topic splitting into account, and the fifth line was obtained also taking the weather forecast merging into account. The next two lines used 3 shows of the RTP07 corpus (RTP07-3) and show the improvements achieved with the integration of non-news information (without and with, respectively). The following two lines used 6 shows of the RTP07 corpus, and show the improvements achieved with the integration of the ASR results (without and with, respectively). The last two lines of the Table show the results that would be achieved if the evaluation window is extended to 2s.

Our collaboration with video segmentation experts in the framework of European project VIDI-VIDEO and a preliminary experiment with a single recent BN allows us to discuss the feasibility of using video derived cues for the task of TS. The fusion of our topic segmentation boundaries derived only from the audio signal with the ones provided by a shot segmentation module may contribute towards a higher precision of the automatically computed boundaries. In terms of video shot representation, semantic concepts such as single news anchor, double news-anchor, news studio, etc. may contribute towards making the overall topic segmentation system more robust and autonomous. The detection of a split screen showing both the lead anchor and the field reporter might also be useful since it never happens at the very begining of a story. These are the type of video derived cues we are currently studying for the potential integration with our audio-based TS module.

5 Conclusions

This paper described our on-going work on the TS module for broadcast news. It summarized our first experiments with a single-stage CART based approach, which explored only the typical structure of BN shows. This approach evolved into a multi-stage approach, which allowed more complex structures with thematic anchors and commentators, and later also explored the topic related structure, the non-news information and the automatically produced transcripts of the BN shows.

The performance of the automatic topic segmentation module was compared with the manual segmentation done by a professional media watch company, yielding quite satisfactory results. The paper also discussed how these could be improved by merging with video derived cues, which is part of our current plans.

6 Acknowledgments

The present work is part of Rui Amaral's PhD thesis, initially sponsored by a FCT scholarship. This work was partially funded by PRIME National Project TECNOVOZ number 03/165, and by the European project Vidi-Video. The authors would like to acknowledge the continuing support of our colleagues J. Neto, H. Meinedo, and V. Mezaris.

References

1. Neto, J., Meinedo, H., Amaral, R., Trancoso, I.: A system for selective dissemination of multimedia information resulting from the alert project. In: Proc. MSDR '2003, Hong Kong (April 2003)
2. Barzilay, R., Collins, M., Hirschberg, J., Whittaker, S.: The rules behind roles: Identifying speaker role in radio broadcast. In: Proc. AAAI 2000, Austin, USA (July 2000)
3. Amaral, R., Trancoso, I.: Exploring the structure of broadcast news for topic segmentation. In: Proc. LTC'07, Poznan, Poland (October 2007)