Topic Development Based Refinement of Audio-Segmented Television News

Alfredo Favenza[1], Mario Cataldi[1], Maria Luisa Sapino[1], and Alberto Messina[2]

[1] Universita' di Torino,Italy
(alfredofa@tiscali.it, {cataldi,mlsapino}@di.unito.it)
[2] RAI - Centre for Research and Technological Innovation, Torino, Italy
(a.messina@rai.it)

Abstract. With the advent of the cable based television model, there is an emerging requirement for random access capabilities, from a variety of media channels, such as smart terminals and Internet. Random access to the information within a newscast program requires appropriate segmentation of the news. We present text analysis based techniques on the transcript of the news, to refine the automatic audio-visual segmentation. We present the effectiveness of applying the text segmentation algorithm CUTS to the news segmentation domain. We propose two extensions to the algorithm, and show their impacts through an initial evaluation.

1 Introduction and related work

For television program repurposing applications, it is important to be able to combine and reuse fragments of existing programs. In order to identify and combine fragments of interest, the editing author needs instruments to query, and efficiently retrieve relevant fragments. Text analysis techniques can support and refine the segmentation results returned by the audio/visual processing methods. Video segmentation has been a hot research topic since many years [4]. The common goal of most technical approaches is to define automatic techniques able to detect the *editorial parts* of a content item, i.e. modal-temporal segments of the object representing a semantic consistent part *from the perspective of its author* [8]. Very recent approaches achieve this goal with interesting results [3]. Significant results in the area of Digital Libraries have been achieved by the partners of the Delos European project [1].

In this paper, we address the problem of segmenting the television news in self contained fragments, to identify the points in which there is a significant change in the topic of discussion. Different units will be indexed for efficient query retrieval and reuse of the digital material. We experiment our method on data available in the archives of RAI CRIT (Center for Research and Technological Innovation) in Torino. In particular, we analyze the transcript of the news, with the goal of combining the resulting text-based segments with the ones extracted by algorithms based on the visual features of the news programs.

In the next sections we first describe the text segmentation algorithm, *CUTS (CUrvature based development pattern analysis and segmentation blogs and other Text Streams)* [7], which focusses on the segmentation of texts of different nature (originally, blog

files) with the guidance of the information about how the addressed topics evolve along time. Our choice of CUTS, as opposed to text-tiling [5], or to the approaches based on the detection of minima in text similarity curves [6, 2] is based on the specific application domain we are dealing with. We apply CUTS on the spoken text, as it is extracted by an Automatic Speech Recognition engine. We then extend the original CUTS method to take into account multi-dimensional curves, and the actual temporal duration of the entries. We discuss our initial experimental results, which show that the extended versions improve on the precision of the segmentation technique within the context of news transcript segmentation.

2 Background: CUTS algorithm

In [7], authors propose CUTS, curvature based development pattern analysis and segmentation for blogs and other text streams. Given a sequences of ordered blog entries, CUTS captures the topic development patterns in the sequence, and identifies coherent segments as opposed to sequences in which the discussion is smoothly drifting from one topic to another, or a main subject is interrupted. There are three main phases in CUTS algorithm.

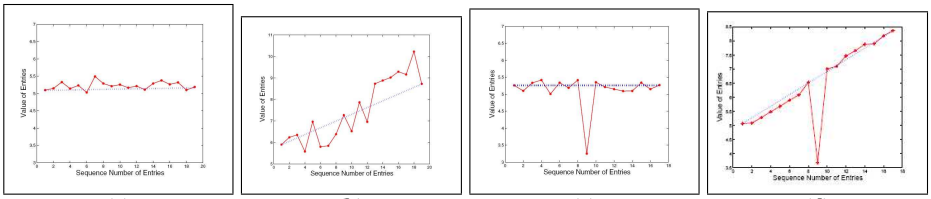
(i) First, the sequence of entries is analyzed, and a representative surrogate (a keyword vector) is generated for each entry. The N text entries are represented in the standard TF/IDF form. As usual, the surrogate generation includes a preliminary phase of stop word elimination and stemming. Each vector component, $w_{k,j}$ represents the weight of the $j-th$ term in the vocabulary w.r.t the $k-th$ entry. Weights are at the basis of the entry-similarity evaluation:

$$s_{i,j} = \sum_{k=1}^{n} w_{i,k}.w_{j,k}$$

The topic evolution is actually computed by referring to the *dissimilarity* among the entries. Pairwise entries dissimilarity is stored in a dissimilarity matrix, D, whose elements $D_{i,j} = 1 - s_{i,j}$, represent the degree of dissimilarity between the $i-th$ and the $j-th$ entries.

(ii) The sequence of entries is then mapped onto a curve, which highlights the development patterns in terms of the (dis)similarity between adjacent entries. CUTS maps the initial data in a new, 1-dimensional space by applying multidimensional scaling algorithm [9]. The mapping preserves to the best the distances between the points. A second dimension plays the role of a temporal dimension[3]. In the resulting space, the consecutive entries form a curve (referred to as the CUTS-curve). By analyzing this curve, in [7] the authors identify different development patterns, illustrated in figure 1. *Dominated* segments are characterized by a sort of "stability" of the topic, which in the CUTS curve, corresponds to an almost horizontal segment. In *Drifting* patterns a smooth transition from a topic to another is observed. As the CUTS curve reflects the differences between consecutive entries, this topic development reflects in a slope in a sloping the curve. The slop of the curve measures how fast one topic evolves in the

[3] In section 3.2 we will show how the final segmentation can benefit from a different dimensional choice.



(a) (b) (c) (d)

Fig. 1. Example of (a) dominated curve segment (b) drifting curve segment (c) interrupted dominated curve segment and (d) interrupted drifting curve segment

following one. *Interrupted* patterns are detected when a sudden, significant change occurs. An interruption is seen as a new topic, which is addressed for a very short amount of time (w.r.t. the duration of the other topics). Interruptions may occur both within dominated and within drifting segments. In audio-segmented text, due to low quality of segmentation, interrupts are common and need to be recognized as such for effective news topic identification. This is one of the reasons we chose CUTS over other schemes.

(iii) The pattern development curve is then analyzed, and the topic segments, as reflected by the changes in the slopes of the curves, are identified. The algorithm introduces an iterative method to represent the curve with a series of straight line segments, each denoted by a $4-tuple\ g_i = (k_i, \sigma_i, (x_s, y_s)_i, (x_e, y_e)_i)$, where (i) k_i is the slope of the segment, and reflects the speed of the topic change in the segment; (ii) σ_i measures the concentration of the entries around the dominant development pattern, and reflects how well the segment approximates the original curve. Good approximation reflects in a small value for σ_i, which is the average of the distances from the original points to the line segment; (iii) $(x_s, y_s)_i$, and $(x_e, y_e)_i$ denote the first and the last points of the segment, respectively.

The decision on whether two consecutive topic segments should be combined in the same topic segment is based on the comparison of their corresponding curve parameters, to measure how homogeneous the segments are in terms of their slope and concentration. Two segments are considered homogeneous (and thus combinable) if

$\mid k_i - ki + 1 \mid < \lambda_{drifting}$ and $\mid \sigma_i - \sigma_{i+1} \mid < (\sigma_i + \sigma_{i+1})/2$

The new parameter $\lambda_{drifting}$ is a threshold to determine when two different topic evolution speed can be considered as homogeneous. It is described in terms of the difference of slopes of the curve segments. After two curve segments are combined, the k and σ values of the resulting segment are calculated. Then the combination process is iteratively repeated.

The annotation process associates each base topic segment with a label characterizing its topic development pattern as *dominated*, *drifting*, or *interrupted*.

3 Extending CUTS for news segmentation

Using a naive application of CUTS to the domain of the news we had misalignment in terms of information loss due to the choice of (i) the dimensionality reduction, scaling the initial multidimensional domain to a monodimensional one, and (ii) the uniform treatment of the duration of the entries.

To address these issues, we propose two extensions to the original CUTS algorithm. The first one (Section 3.1) extends the method to the 3-dimensional case, in which the first two dimensions are returned by the multidimensional scaling method, and the third one is - as in the original method - the positional temporal dimension. The second extension (Section 3.2) takes into account the actual duration of the entries.Finally, we discuss the impact of the combination of the two extensions, returning a 3-dimensional, temporal approach.

3.1 Multi-dimensional CUTS

In our first extension, *the sequence of news entries is represented by curves in a 3-dimensional space*. The x-axis will be associated with the temporal dimension (mapping consecutive entries to consecutive unit time instants). The y and z axes are associated to the dimensions returned by the 2-dimensional scaling. While the main idea of the CUTS algorithm remains unchanged, the extension affects the way distances, breakpoints, drifting, and all the parameters (segments slopes, segment composition conditions) needed to analyze the 3-dimensional curves, are defined.

In the original CUTS method [7], the choice of the break points to identify segments is based on the distance between pairs of points and lines, both of them in the two dimensional space. A similar strategy is adopted in the 3-dimensional space, of course by referring to the appropriate methods to evaluate the distances between a point and a segment in the 3-dimensional space.

Generalization of curve model to n-dimensional space The drifting condition can be expressed as a condition on the difference between two consecutive entries of $\hat{J_{S(t)}}$, i.e. the finite difference of the Jacobian vector of the multidimensional curve.

$$\| \Delta \hat{J_{S(t)}} \|^2 \leq \Lambda^2 \tag{1}$$

being Λ the discriminating threshold. Under the hypothesis of unit duration associated to the entries, the intervals of duration of two consecutive entries are $[i, i+1]$, and $[i+1, i+2]$, respectively, and their duration is, by hypothesis, exactly 1. Thus

$$\| \Delta \hat{J_{S(t)}} \|^2 = \sum_{i=0}^{n} [(d_i(i+2) - d_i(i+1)) - (d_i(i+1) - d_i(i))]^2$$

In general, a possible way to estimate the value for the parameter Λ, when dealing with two dimensions, is to compute the global variation of the similarity along the two corresponding axes. For any dimension d_i, we will have $\lambda_i = \alpha_i \frac{\delta d_i}{\delta t}$, where δd_i is the excursion of d_i, δt is the corresponding excursion of t, and α_i is the weight of the variation along the $i-th$ direction. The above condition (1) is thus rewritten as

$$\| \Delta \hat{J_{S(t)}} \|^2 \leq \sum_{i=0}^{n} \lambda_i^2$$

Notice that this is a general formulation of the drifting condition, which, in the case $n = 1$, and entries of duration 1, reduces to the condition applied by the original CUTS,

$$[(d_1(i+2) - d_1(i+1)) - (d_1(i+1) - d_1(i))]^2 \leq \alpha \delta d_1$$

that is,

$$| k_{i+1} - k_i | \leq \lambda_{drifting}$$

Seg	Base S.	Sec.	Ann	Seg	Base S.	Sec.	Ann
1	0–3	0–34	drif	12	60–62	892–921	dom
2	3–4	34–61	drif	13	62–63	921–937	dom
3	4–26	61–383	dom	14	63–73	937–1064	dom
4	26–28	383–415	dom	15	73–86	1064–1269	dom
5	28–37	415–502	drif	16	86–98	1269–1469	dom
6	37–38	502–517	dom	17	98–100	1469–1503	dom
7	38–52	517–755	dom	18	100–103	1503–1535	dom
8	52–53	755–773	dom	19	103–112	1535–1669	dom
9	53–54	773–780	drif	20	112–119	1669–1794	dom
10	54–59	780–872	drif	21	119–120	1794–1806	dom
11	59–60	872–892	dom				

Seg	Base S.	Sec.	Ann
1	0–3	0–34	drif
2	3–4	34–61	drif
3	4–53	61–773	dom
4	53–59	773–872	inter
5	59–108	872–1595	dom
6	108–109	1595–1614	drif
7	109–110	1614–1632	drif
8	110–120	1632–1806	dom

Table 1. Automatic monodimensional and two-dimensional temporal CUTS based annotation .

3.2 Extensions to the temporal dimension

Our input text fragments also contain information about their actual temporal duration within the news programme.

Technically, taking into account the temporal dimension only requires the actual instantiations of the time intervals appearing in the general definition of the function $\| \Delta \hat{J_{S(t)}} \|^2$ in Section 3.1. In fact, in Section 3.1 we introduced simplified expressions for the function, reflecting the fact that all the entries were associated to a time unit duration. In this case, denoting with $[st_i, et_i]$ and $[st_{i+1}, et_{i+1}]$ the time intervals associated to the $i-th$ and the $(i+1)-th$ entries respectively[4], the corresponding expression for $\| \Delta \hat{J_{S(t)}} \|^2$ becomes

$$\| \Delta \hat{J_{S(t)}} \|^2 = \sum_{i=0}^{n} \left(\frac{d_i(et_{i+1}) - d_i(st_{i+1})}{et_{i+1} - st_{i+1}} - \frac{d_i(et_i) - d_i(st_i)}{et_i - st_i} \right)^2$$

Table 1 reports the results of the segmentation and annotation methods discussed in the previous sections, extended with the temporal dimension.

We notice that in both cases we see that extensions improve the qualities of the results, giving a closer approximation of the human domain expert classification.

4 Analysis of results

To evaluate the methods, we measure the coherence of the automatically detected topic segment boundaries (S_1) wrt, the segment boundaries in the ground truth (S), i.e., a segmentation given by a human domain expert. We define *precision* as the ratio of entry numbers from the automatically extracted sequence which are common boarder list ($Prec = (entry(S_1) \cap entry(S))/(entry(S_1))$), and recall as the ratio of entries which delimit topic segments in the ground truth, and are also returned by the automatic system ($Rec = (entry(S_1) \cap entry(S))/(entry(S))$. We also notice that the above measure penalizes the cases in which the disagreement between the ground truth and the returned sequence of topic segments is small (for example, the cases in which there is a one entry displacement between the automatically detected topics and the ground

[4] In this specific application it holds that $et_i = st_{i+1}$.

truth). To take into account the displacement, we define an embedding procedure which aligns the two topic sequences to be compared, and we compute the alignment cost as follows: $Align\_cost = d + \sum_i((s_i - 1) + (s_i \times da_i))$, where, d is the summation of the differences between the i-th element, $e1_i$ of the shorter sequence and the closer element, $e2_i$ appearing in the other sequence of entries, s_i is the number of segments between between $e2_i$ and $e2_{i-1}$, i.e., the number of segments in the second sequence that all together correspond to a single segment in the first one, and da_i captures the highest disagreement in the annotation between the $i - th$ segment in the first sequence and any of the corresponding s_i segments, computed by measuring the positional distance between the annotations, wrt.the ordering $dominated \prec drifting \prec interrupted$.

Table 2 shows the results on the example presented in the previous sections.

	2MDS	**2MDS + Time**	**1MDS**	**1MDS + Time**
Precision	0.681818	0.709415	0.271916	0.295726
Recall	0.616666	0.645833	0.205349	0.270398
Alignment cost	28	17	67	56

Table 2. Statistical evalutation of results

5 Conclusions and future work

We have presented a text analysis based technique on the transcript of the news, to refine the automatic audio-visual segmentation. In particular, we have discussed and evaluated the effectiveness of applying the text segmentation algorithm CUTS [7], and our two extensions of it, to the news segmentation domain. The initial evaluation results are consistent with our expectations. We are now conducting intensive user studies to evaluate the methods on different data and users samples. We are also working on the integration of the text based segmentation results with the segments extracted through audio-visual methods.

References

1. Delos network of excellence on digital libraries. Internet Site http://www.delos.info/.
2. P. Andrews. Semantic topic extraction and segmentation for efficient document visualization. Master's thesis, School of Computer & Communication Sciences, Swiss Federal Institute of Technology, Lausanne, 2004.
3. S. Venkatesh D.-Q. Phung, T.-V. Duong and H. H. Bui. Topic transition detection using hierarchical hidden markov and semimarkov models. In *Proc. of ACM Multimedia 2005*, 2005.
4. A. Smeaton H. Lee and N.E. O'Connor. User evaluation of físchlár-news: An automatic broadcast news delivery system. *ACM Transactions on Information Systems*, 24(2):145–189, 2006.
5. Marti A. Hearst. Texttiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics*, 23(1):33–64, 1997.
6. Marti A. Hearst and Christian Plaunt. Subtopic structuring for full-length document access. In *Proc. of SIGIR*, 1993.
7. Y. Qi and K.-S. Candan. Cuts: Curvature-based development pattern analysis and segmentation for blogs and other text streams. In *Proc. of Hypertext 2006*, 2006.
8. C. G. Snoek and M. Worring. Multimodal video indexing: A review of the state-of-the-art. In *Proc. Multimedia Tools and Applications*, pages 5–35, 2005.
9. W. S. Torgerson. Multidimensional scaling: I. theory and method. *Psychometrika*, 17(4), 1952.