HITS and IRISA at MediaEval 2013: Search and Hyperlinking Task

Camille Guinaudeau
HITS
Schloss-Wolfsbrunnenweg 35
D-69118 Heidelberg, Germany
firstname.lastname@h-its.org

Anca-Roxana Simon*, Guillaume Gravier**, Pascale Sébillot***
IRISA & INRIA Rennes
Univ. Rennes 1*, CNRS**, INSA***
35042 Rennes Cedex, France
firstname.lastname@irisa.fr

ABSTRACT

This paper describes our approach and results in the hyperlinking sub-task at MediaEval 2013. A two step method is implemented where the first step consists in establishing a shortlist of relevant videos. In the second step, a target segment is selected from each video in the shortlist. We focus on target selection comparing two distinct strategies. The first one exploits a bipartite graph relating utterances and words to find the most relevant utterances from which segments are derived. The second one uses explicit topic segmentation, whether hierarchical or not, to select the target segments.

1. INTRODUCTION

We present the joint participation of HITS and IRISA to the Search and Hyperlinking task at MediaEval 2013 [2], limiting ourselves to the hyperlinking sub-task where one is required to find targets for hyperlinks whose source is a given anchor. Similar to last year, we adopt a two step approach. A shortlist of semantically related target videos is first established by comparing the anchor, possibly with context, to entire videos using standard information retrieval techniques. In the second step, we search for the most relevant target segment within each video in the shortlist, respecting the time constraints imposed.

In 2013, we focused on the last step, i.e., the selection of the most relevant target segment inside each video in the shortlist of semantically related videos. We believe that precise target selection is a crucial step for the hyperlinking task: wrong timestamps within semantically related videos can make the result useless even though the video is per se relevant. However, previous work on the hyperlinking sub-task [1] mostly focused on linking anchors with relevant videos but did not pay much attention to precise target selection. We implemented two distinct approaches of which several variants are compared. The first approach relies on a link analysis algorithm which exploits links in a graph to propagate associations between words and utterances so as to select a small number of utterances as the link target. The second one relies on explicit topic segmentation to find out topically coherent targets closely related to the anchor, extending last year's approach to hierarchical segmentation and fine grain text alignment techniques.

2. SYSTEM DESCRIPTION

We mostly exploit the transcripts provided [3, 6], which were lemmatized, keeping only nouns, non modal verbs and adjectives. Utterances are sentences for manual transcripts, speech segments for LIMSI's and shots for LIUM's[1].

2.1 Shortlist of semantically related videos

To limit detailed search for hyperlink targets given an anchor, we first establish a shortlist of the 50 most related videos, considering each video as a whole. A vectorial representation of transcripts is used for both anchors and videos, adopting the BM25 weighting. When the context in which the anchor appears is considered, a linear combination of the BM25 weights obtained resp. from the anchor and from the context is used, with a strong emphasis on the anchor (0.8 *vs.* 0.2). Videos are ranked in decreasing order according to the cosine distance with the anchor (possibly with its context), removing videos which contain the anchor (same file or file corresponding to rebroadcasting of the same content). The shortlist contains the top 50 videos to which we want to relate the anchor and which are further processed to select a precise and short enough hyperlink target.

2.2 Selection of hyperlinks targets

For each item in the top 50 related videos, we need to extract the target segment for the link that will be established with the anchor. According to evaluation rules, target should be an excerpt with a duration between 10 s and 2 min. Two approaches were taken, based on the same underlying idea, i.e., finding the consecutive shots or utterances within the given time constraints which are the most related to the anchor. A first approach relies on the hyperlink-induced topic search (HITS) algorithm [5], a link analysis method used to weight each shot according to its relationships with words from the anchor. A second approach implements topic segmentation to find out coherent segments which are compared to the anchor.

Target selection with link analysis.

For a given shortlist video, link analysis relies on a bipartite graph where the first set of nodes represents utterances, the second one representing words. Edges reflect the pairing between words and utterances, i.e., an edge between utterance S_i and word W_j indicates that W_j appears in S_i.

[1]Utterance boundaries being absent from LIUM's transcripts, alignment with shot boundaries was performed.

Exploiting the bipartite graph structure, the HITS algorithm aims at assigning a score to each node n in the graph, where the score indicates how well n is connected to the others. HITS iteratively propagates scores via edges, taking into account the importance of nodes connected to n. In the framework of hyperlink target selection, the idea is to give a high score to utterances that are connected to words related to the anchor and its context. Scores in word nodes are initialized with a value reflecting the word frequency in the anchor, alone (**HITS**^{a}) or with context (**HITS**^{c}). Frequent words increase the score of utterances containing such words, in turn improving the score of words that appear in the vicinity (i.e., the same utterance) of anchor words.

After convergence of the HITS algorithm, a score is obtained for each shot by adding the scores of all utterances within the shot. Merging heuristics are finally used to yield segments from which the best scoring one is picked as the link target. Adjacent shots with a score above a threshold are merged into a single segment if the result is less than 2 min long, adding scores. Short segments less than 10 s are merged with the highest scoring neighbor.

Target selection with topic segmentation.

As an alternative to link analysis, linear and hierarchical topic segmentation is used to partition each video in the shortlist into homogeneous segments. Each segment is compared to the anchor, considered with its context in all topic segmentation experiments, to find the most significant one.

Linear topic segmentation is achieved using [4], providing a set of segments which exhibit high vocabulary coherence. In the hierarchical approach, each segment resulting from linear segmentation is again segmented using a criteria which combines lexical cohesion and disruption [7] so as to avoid over-segmentation. The idea of hierarchical segmentation is to have smaller segments to relate to the anchor, thus possibly more accurate targets.

For each segment resulting either from linear or from hierarchical topic segmentation, the similarity with the anchor and its context is calculated. We investigate two distances. The first one is a classical cosine similarity measure assuming *tf-idf* weights, thus relying on a bag of words representation. This strategy was applied to linear (**Linear+BoW**) and to hierarchical segmentation (**Hierarchical+BoW**). To achieve better comparison, we also experimented n-gram alignments, where similarity is computed between words, bigrams and trigrams separately. Similarities from different n-gram orders are linearly combined with weights equal to 0.2, 0.3 and 0.5 for order 1, 2 and 3 respectively. N-gram comparison was applied to linear segmentation (**Linear+ngrams**).

The best scoring segment is used as target, applying the following postprocessing rules to match time constraints. Segments longer than 2 min are resegmented using a sliding window of 2 min, taking the best scoring window within the segment. Segments shorter than 10 s are combined with the best scoring neighbor until the minimum length is reached.

3. RESULTS

A number of observations can be drawn from the official evaluation results in Tab. 1.

Considering the anchor and its context, the best results are clearly obtained with n-gram alignment along with linear topic segmentation. These good results are obviously to be attributed to n-grams which yields target segments whose content is closely related to that of the anchor, if not almost similar. This tends to indicate that evaluators on Amazon Mechanical Turk (AMT) prefer links to highly correlated content as opposed to links targeting contents on the same subject but with a more remote relationship.

	LIMSI	LIUM	MANUAL
HITS^a	0.0328	0.0253	—
HITS^c	0.0305	0.0237	—
Linear+BoW	0.0219	0.0281	0.0436
Linear+ngrams	0.0399	0.0467	0.0633
Hierarchical+BoW	0.0193	0.0233	0.0362

Table 1: Results for all methods on the 2013 test set

Hierarchical segmentation turned out to be deceiving. One probable explanation is that targets are somewhat smaller than for linear segmentation (half the length of segments obtained using linear segmentation on average). Small target segments make comparison with the anchor less reliable and increase the probability of having poorly related content.

Using HITS as described in Sec. 2.2 appears as a good strategy for target selection. HITS implicitely uses a bag of words representation and compares favorably with linear topic segmentation when comparison with the anchor relies on a similar representation. Introducing n-grams in the graph might be a good option to improve the HITS-based approach.

Finally, topic segmentation algorithms yield better results on the LIUM transcripts than on LIMSI transcripts. This is most likely due to the fact that utterances in LIUM transcripts correspond to visual shots. Hence, the resulting target is visually consistent, while this is not the case for LIMSI transcripts when using topic segmentation which relies on utterances that are not related to visual content (LIMSI's utterances are usually longer while reference utterances are smaller). We believe that visual consistency is a crucial factor for AMT evaluators.

4. REFERENCES

[1] M. Eskevich, G. J. F. Jones, R. Aly, and et al. Multimedia information seeking through search and hyperlinking. In *ACM Intl. Conf. on Multimedia Retrieval*, 2013.

[2] M. Eskevich, G. J. F. Jones, S. Chen, R. Aly, and R. Ordelman. The Search and Hyperlinking task at MediaEval 2013. In *Working notes of the MediaEval 2013 Workshop*, 2013.

[3] J.-L. Gauvain, L. Lamel, and G. Adda. The LIMSI broadcast news transcription system. *Speech Communication*, 37(1-2):89–108, 2002.

[4] C. Guinaudeau, G. Gravier, and P. Sébillot. Enhancing lexical cohesion measure with confidence measures, semantic relations and language model interpolation for multimedia spoken content topic segmentation. *Computer Speech and Language*, 26(2):90–104, 2011.

[5] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999.

[6] H. Schwenk and P. Lambert. LIUM's SMT machine translation systems for WMT 2011. In *Workshop on Statistical Machine Translation*, 2011.

[7] A. Simon, G. Gravier, and P. Sébillot. Leveraging lexical cohesion and disruption for topic segmentation. In *Empirical Methods in NLP*, 2013.