Searching and Hyperlinking using Word Importance Segment Boundaries in MediaEval 2013

Kim Schouten
Erasmus University Rotterdam
Rotterdam, the Netherlands
schouten@ese.eur.nl

Robin Aly
University of Twente
Enschede, the Netherlands
r.aly@ewi.utwente.nl

Roeland Ordelman
University of Twente
Enschede, the Netherlands
r.aly@ewi.utwente.nl

ABSTRACT

This paper reports a set of experiments performed in the context of the *Searching and Hyperlinking* task of the MediaEval Benchmark Initiative 2013. The *Searching* part challenges to return a ranked list of video segments that are relevant given some textual user query, while for the *Hyperlinking* task the aim is to return a ranked list of video segments that are relevant given some video segment. The main focus is on finding a way to compute flexible segment boundaries. This is performed by extending the term frequency part of tf-idf to include the temporal dimension of videos. Although the contribution is theoretically sound its performance is relatively poor, which we attribute to the focus on speech data and the hyperlinking process. We plan to refine our method in the future overcome these limitations.

1. INTRODUCTION

The content of videos can be long and current search approaches that return whole videos waste the time of users searching for specific information. The Search and Hyperlinking Tasks in the MediaEval Benchmark Initiative 2013 [3], which is a refinement of its previous instance [1], models the situation where users should be directly pointed to relevant segments within videos and be offered links to video segments relevant to the already found video segment. State-of-the-art search methods use fixed video segments as a return unit. However, fixed video segments have the limitation that evidence for a relevant passage can be divided between two segments or that the segments are too long. Therefore, in this paper we propose a method to determine segment boundaries that are suitable for individual queries and to rank these segments accordingly.

The paper is outlined as follows. Section 2 describes our methods for search and hyperlinking video. Section 3 describes the details of our experiment and show the evaluation results of our submitted runs, and finally we conclude in Section 4.

2. METHOD

In the following we describe our methods used for the searching and hyperlinking sub-tasks. To this end, we extend the well-known tf-idf ranking function to incorporate the temporal dimension of videos when computing scores. Our intuition is that term are not only important to the moment they were uttered but also to a time window around the utterance. We model the posterior probability of a query term q being important at time t given its utterance at time t_2 as a double sigmoid function:

$$p(q_t = 1|o_t) = \frac{1}{1+\exp(-\frac{t-b+c}{s})} - \frac{1}{1+\exp(-\frac{t-e-c}{s})} \tag{1}$$

where
t = video time in milliseconds
q_t = binary variable of q being important at time t
o_q = the utterance of term q (milliseconds)
b = the begin time of query term q being uttered
e = the end time of query term q being uttered
c = a correction term to shift the sigmoid to, respectively, the left and the right, such that $b < \forall t < e$ yields a probability close to 1
s = the steepness of the sigmoid modeling the reach of the importance around o_q, this is a multiple of $idf(q)$

Given our probabilistic model of the importance of query terms, we now describe how we integrate this model into the tf-idf ranking scheme. Instead of term frequencies in a document, we calculate the score of time t as the frequency of utterances that are important to t times its idf factor. However, because we are not certain about the actual importance q_t, we calculate the expected frequency of important terms:

$$score(t) = \sum_{q\in Q}\sum_{o_q\in O_q} idf_q E[q_t|o_q] = \sum_{q} idf_q \sum_{o_q\in O_q} p(q_t|o_t) \tag{2}$$

where Q are the query terms and O_q are all occurrence of term q in the video.

Based on this ranking scheme, we determine the extent of a result segment s as the adjacent t's that have a score above a certain threshold (set at 0.01 in our experiments). The score of a segment seg is set to be the maximum $score(t)$ for every time point t in seg (we break ties by selecting the shorter segment):

$$score(seg) = \max_{t\in seg} score(t) \tag{3}$$

where t iterates over the time interval of the segment. Figure 1 shows an example of our ranking scheme using a term with large idf and a term with low idf with each one utterance, as well as the score(t) function that combines both.

MediaEval 2013 Workshop, October 18-19, 2013, Barcelona, Spain

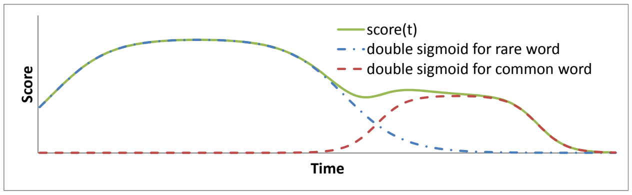

Figure 1: An example plot of two query keyword occurrences in a video.

For the hyperlinking task, we construct a textual query from the given anchor segment. To this end, we consider the words uttered during the anchor segment and select the ten words with the largest Kullback-Leibner divergence compared to the language model of the whole collection, loosely following the method proposed in [4]. Using the query, we used the same ranking scheme as for the search task.

3. EXPERIMENTS

In this section, we describe the experimental setup and submitted runs for the searching, and hyperlinking sub-tasks. For both search queries and constructed hyperlinking queries, we return a ranked list of video segments by decreasing score.

3.1 Setup

For an overview of the used dataset please refer to [3]. Transcripts from LIUM are provided in blocks of 20 seconds, which we chose as the duration of each term. For the manual subtitles and LIMSI transcripts we used the time annotation for the individual word.

Our ranking scheme requires two parameters, see Equation 1. We set $c = s * \log(\frac{1.0}{0.01} - 1)$, which ensures that the sigmoids will yield a value of 0.99 between b and e. This corresponds to the intuition that a term is almost certainly important close to the time it was uttered. For the steepness parameter, we use $s = 100,000 * idf(q)$, causing rare words to have wider influence.

3.2 Results

The results for the searching sub-task are given below in terms of Mean Reciprocal Rank (MRR), mean Generalized Average Precision (mGAP), and Mean Average Segment Precision (MASP) (cf. [2]).

Runs	MRR	mGAP	MASP
LIUM	0.09	0.06	0.06
LIMSI	0.08	0.04	0.04
BBC Subtitles	0.05	0.04	0.03

For the hyperlinking sub-task, the results are reported in terms of Mean Average Precision (MAP).

Runs	MAP
LIUM	0.4081
LIMSI	0.4048
BBC Subtitles	0.3120

We see that the subtitles perform the poorest in both task. The LIUM transcripts outperform the ones from LIMSI, albeit with only small differences, which were statistical insignificant ($p = 0.05$). In general, these results are in line with our previous work [5].

3.3 Discussion

The described method shows poor performance. We identify the following possible reasons: First, our ranking scheme only uses speech information to find video segments, while their relevance is sometimes determined by their visual content or in the metadata (we found several instances where this is the case). Second, we used only one parameter setting, which was the most intuitive to us. We believe other settings can improve the performance. Finally, for the query generation for the hyperlinking task, we only used the text uttered during the anchor segments. As some of them are relatively short (with a minimum of 10sec), we believe considering their surrounding can improve their performance.

4. CONCLUSIONS

We have described a method to rank video segments based on the uncertain importance of query terms at particular time in a video given the term's utterances in the transcript. The time interval where a term is important was described by a probability distribution modeled as a double sigmoid function. We proposed that the points where the probability that none of the terms is important are intuitive segment boundaries specific to the current query, which many methods lack. To rank segments we expanded the standard tf-idf weighting scheme to the situation where the importance of query terms at a given was uncertain. The final score to rank segments was the maximum expected tf-idf score.

The described method showed relatively poor performance. We plan to pursue the following paths to improve the results in the future. First, we plan to extend our scheme to incorporate visual content and metadata. Second, we will investigate multiple parameter settings to tune our method. Finally, for hyperlinking we only extracted words from the given segment, which we believe may not provide enough information. In the future we plan to include the surrounding of the anchor segment and the initial query for which a user arrived at the current segment.

5. REFERENCES

[1] M. Eskevich, G. J. F. Jones, S. Chen, R. Aly, R. Ordelman, and M. Larson. Search and Hyperlinking Task at MediaEval 2012. In *MediaEval 2012 Workshop*, Pisa, Italy, October 4-5 2012.

[2] M. Eskevich, W. Magdy, and G. J. F. Jones. New metrics for meaningful evaluation of informally structured speech retrieval. In *Proceedings of the 34th European conference on Advances in Information Retrieval*, ECIR'12, pages 170–181. Springer Berlin Heidelberg, 2012.

[3] Maria Eskevich, Gareth J.F. Jones, Shu Chen, Robin Aly, and Roeland Ordelman. The Search and Hyperlinking Task at MediaEval 2013. In *MediaEval 2013 Workshop*, Barcelona, Spain, October 18-19 2013.

[4] V. Lavrenko and W. Bruce Croft. *Language Modeling for Information Retrieval*, chapter Relevance models in information retrieval, pages 11–56. Kluwer Academic Publishers,, 2003.

[5] D. Nadeem, R. Aly, and R. Ordelman. UTwente does brave new tasks for mediaeval 2012: Searching and hyperlinking. In *MediaEval 2012 Multimedia Benchmark Workshop*, volume 927 of *CEUR Workshop Proceedings*. CEUR, 2012.