CUNI at MediaEval 2013 Search and Hyperlinking Task

Petra Galuščáková and Pavel Pecina
Charles University in Prague
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics
Prague, Czech Republic
{galuscakova,pecina}@ufal.mff.cuni.cz

ABSTRACT

We describe our approach to the Search Subtask of the Search and Hyperlinking Task at MediaEval 2013. We experiment with various methods for segmentation of the recordings into shorter segments which are then used in a standard retrieval setup to search for relevant passages. We use regular segmentation into equilong segments and experiment with machine-learning based segmentation expanding our approach to Similar Segments in Social Spech Task [4].

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: H.3.1 Content Analysis and Indexing; H.3.3 Information search and Retrieval; H.3.4 Systems and Software

1. INTRODUCTION

The main focus of the Search and Hyperlinking Task is to enable users to find information relevant to the submitted query in a collection of audio-visual recordings (Search Subtask) and to find more segments similar to the retrieved ones to enable easier navigation in the collection (Hyperlinking Subtask). The collection consists of TV programmes provided by BBC. The video recordings, audio track, metadata, synopsis, cast, detected shots, detected faces, visual concepts, subtitles and two automatic (ASR) transcripts – provided by LIMSI [5] and LIUM [7] are available. For the Search Subtask, 4 training and 50 test queries are available. More details of the task and the data collection can be found in the task description [2].

2. APPROACH DESCRIPTION

We participated only in the Search Subtask of the Search and Hyperlinking task. In our approach, we divided the recordings into shorter segments and applied the Terrier IR system[1] to search for relevant passages within the recordings. As we searched in the set of the known-boundary segments, we were able to determine the beginnings and ends of the retrieved segments.

As the training set only consists of four queries, we collected other 30 queries and used the whole set consisting of 34 queries for training. We randomly selected recordings from the collection, identified short passages somehow interesting to us in these recordings, and formulated the queries to search for those passages. We formulated the queries to imitate the formulation of the original given queries (e.g. "how to prepare Vietnamese spring rolls", "Thomas Tallis signature", and "a difference between a hare and a rabbit").

We applied the Hiemstra Language Model with the parameter set to 0.35 as it achieved good results in our experiments in the Search and Hyperlinking Task in 2012 [3]. Based on the same results, we decided to apply Porter stemming, implicit set of stopwords, and removal of the overlapping segments from the retrieved results.

We also employed metadata information (title, source, variant, description, service name, episode name, and short episode synopsis) and synopsis, which improved the results on the training data. We simply concatenated each segment (in all recordings) with the metadata information available for the corresponding recording. We also tried to use the cast information, but it slightly decreased the results.

Segmentation of the recordings is realized in two ways: regular segmentation and segmentation based on machine learning (ML).

2.1 Regular segmentation

The regular segmentation achieved good results in last year's experiments [1], so we decided to apply it in this year's experiments as well. The regular segmentation divides the recordings into 50-seconds-long passages. The shift (and overlap) between the adjacent passages is 25 seconds. We use the same segment length and shift as we use in the Similar Segments in Social Speech task [4] to be able to compare the results of both tasks. Moreover, according to the previous experiments, this length and the shift should achieve good results.

2.2 Machine-learning segmentation

The ML-based segmentation is adopted from our experiments conducted for the Similar Segments in Social Speech Task [4]. It employs classification trees to identify segment beginnings and segment ends. The utilized features include cue words and cue tags, letter cases, length of the silence before the word, division given in transcripts, and the output of the TextTiling algorithm [6]. The cue words mainly consist of words which stand frequently at the boundaries. We identified them based on the human transcripts of the recordings in the Similar Segments in Social Speech Task.

As the training set of the Search Subtask is very small (even after the additional training data is included), we de-

[1] `http://terrier.org`

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

Transcripts	Segmentation	Metadata	MRR	MRR-Full	mGAP	MASP
Subtitles	Regular	—	0.285	0.603	0.170	0.197
LIMSI	Regular	—	0.196	0.408	0.108	0.107
LIUM	Regular	—	0.205	0.447	0.124	0.126
Subtitles	Regular	Meta & Syn.	**0.287**	**0.648**	**0.174**	**0.219**
LIMSI	Regular	Meta & Syn.	0.224	0.524	0.112	0.146
LIUM	Regular	Meta & Syn.	0.235	0.570	0.118	0.142
LIUM	ML-Pairs	Meta & Syn.	0.178	0.550	0.106	0.128

Table 1: Retrieval results on the test set for different types of transcripts and segmentation.

cided to use the model trained on the Similar Segments in Social Speech Task.

We at first identified all possible beginnings of the segments and all possible ends of the segments, both tuned for high F-measure. Then, for each possible beginning, we identified the segment end (from the set of possible segment ends) which lay closest to 50.111 seconds (which is the average segment length in the Similar Segments in Social Speech Task) from the beginning.

This approach (when the model is trained on a different data set) enables us to examine the possibility of creating a universal model for ML-based segmentation. However, it also carries potential problems. The sets of the cue words collected on the student dialogues may differ from the cue words used in TV programmes, and the silence between the words in dialogues may have different distribution as the silence between words in the TV programmes.

3. RESULTS

The Search Subtask is evaluated using Mean Reciprocal Rank (MRR), mean Generalized Average Precision (mGAP), and Mean Average Segment Precision (MASP). Details about these measures can be found in the task description [2]. The utilized MRR score considers only points retrieved closer than 60 seconds from the beginning of the segment as correctly retrieved. Therefore, we also utilize the MRR measure in a standard way [8] on full documents, without the window limiting the longest distance between the segment beginning and retrieved beginning – the MRR-Full measure. This measure indicates the quality of the retrieval of the whole recording, the precision of the retrieval of the relevant segment is not taken into account.

Our results for the Search Subtask are presented in the Table 1. The highest result is achieved on the subtitles, using the regular segmentation and employing metadata and synopsis. The subtitles, unsurprisingly, outperform both ASR transcripts in all measures; the LIUM transcripts outperform LIMSI transcripts in all measures, except the MASP score for the regular segmentation if the metadata are employed. The ML-based segmentation is applied on the LIUM transcripts only, and it did not improve the results.

4. CONCLUSIONS

In our approach to the Search Subtask, we employed regular segmentation into 50 seconds long passages on three transcripts. The highest score was achieved on the subtitles. The utilization of metadata slightly improved the results in all applied measures.

We also proposed another solution which interlinks this task and our solution used in the Similar Segment in Social Speech Task and use models trained in the Similar Segment in Social Speech Task to segment the transcripts in the Search and Hyperlinking Task. However, this approach did not bring any positive results and needs to be investigated in more depth.

5. ACKNOWLEDGMENTS

This research is supported by the Charles University Grant Agency (GA UK n. 920913) and the Czech Science Foundation (grant n. P103/12/G084).

6. REFERENCES

[1] M. Eskevich, G. J. Jones, R. Aly, R. J. Ordelman, S. Chen, D. Nadeem, C. Guinaudeau, G. Gravier, P. Sébillot, T. de Nies, P. Debevere, R. V. de Walle, P. Galuščáková, P. Pecina, and M. Larson. Multimedia information seeking through search and hyperlinking. In *Proc. of ICMR*, pages 287–294, Dallas, Texas, USA, 2013.

[2] M. Eskevich, G. J. Jones, S. Chen, R. Aly, and R. Ordelman. The Search and Hyperlinking Task at MediaEval 2013. In *MediaEval 2013 Workshop*, Barcelona, Spain, 2013.

[3] P. Galuščáková and P. Pecina. CUNI at MediaEval 2012 Search and Hyperlinking Task. In *Working Notes Proceedings of the MediaEval 2012 Workshop*, volume 927 of *Workshop Proceeding*, Pisa, Italy, 2012. CEUR Workshop Proceedings.

[4] P. Galuščáková and P. Pecina. CUNI at MediaEval 2013 Similar Segments in Social Speech Task. In *MediaEval 2013 Workshop*, Barcelona, Spain, 2013.

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

[6] M. A. Hearst. TextTiling: Segmenting Text into Multi-paragraph Subtopic Passages. *Computational Linguistics*, 23(1):33–64, 1997.

[7] H. Schwenk, P. Lambert, L. Barrault, C. Servan, H. Afli, S. Abdul-Rauf, and K. Shah. LIUM's SMT machine translation systems for WMT 2011. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, WMT '11, pages 464–469, Edinburgh, UK, 2011.

[8] E. M. Voorhees. The TREC-8 Question Answering Track Report. In *Proceedings of TREC-8*, pages 77–82, Gaithersburg, MD, USA, 1999.