DCU Linking Runs at MediaEval 2013: Search and Hyperlinking Task

Shu Chen
INSIGHT Centre for Data Analytics / CNGL
Dublin City University
Dublin 9, Ireland
shu.chen4@mail.dcu.ie

Gareth J. F. Jones
CNGL, School of Computing
Dublin City University
Dublin 9, Ireland
gjones@computing.dcu.ie

Noel E. O'Connor
INSIGHT Centre for Data Analytics
Dublin City University
Dublin 9, Ireland
Noel.OConnor@dcu.ie

ABSTRACT

We describe Dublin City University (DCU)'s participation in the Hyperlinking sub-task of the Search and Hyperlinking of Television Content task at MediaEval 2013. Two methods of video hyperlinking construction are reported: i) using spoken data annotation results to achieve the ranked hyperlink list, ii) linking and merging meaningful named entities in video segments to create hyperlinks. The details of algorithm design and evaluation are presented.

Keywords

Hyperlinking, Multimedia Search, Anchor Selection, Information Retrieval

1. INTRODUCTION

This paper presents Dublin City University (DCU)'s participation in the Hyperlinking sub-task of Search and Hyperlinking of Television Content task at MediaEval 2013. The paper is organized as follows: Section 2 describes our automatic hyperlinking strategies, Section 3 gives experimental results, and Section 4 concludes the paper.

2. HYPERLINKING STRATEGIES

2.1 Hyperlimking Principles

In this subsection we describe the principles underlying our approach to the hyperlinking task. The elements involved in the hyperlinking framework correspond to the query anchor, the target segment, and the hyperlink. The query anchors, as the input to the hyperlinking framework, are defined in [1]. A target segment is a subset of a video to which a query anchor is supposed to be linked. For our approach, a fixed window whose duration is 120 seconds and the overlap is 30 seconds is used to determine the target segments. The spoken data in the video is available in three transcripts: automatic speech recognition (ASR) transcripts from LIUM Research [6], LIMSI/Vocapia [2] and manual subtitles provided by the BBC [1]. Hyperlinks are constructed from the query anchor to a set of target segments using different hyperlinking strategies as described in the following subsections.

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

2.2 Hyperlinking using Text Annotation

This strategy determines the hyperlinks based on two quality measures: the video-level and the segment-level. The video-level measure aims to determine the relevance between the video containing the query anchor and other videos containing potential target segments based on the text transcripts. DBpedia Spotlight[1], implementing text annotation by supervised learning through DBpedia Ontology[2], was used to extract a set of terms to represent the textual content of each video. The method used to annotate terms in DBpedia Spotlight is based on a TF*ICF model [4], where TF (Term Frequency) represents the relevance of a term in the spoken video, and ICF (Inverse Candidate Frequency) is determined by the relevance of a term in DBpedia Ontology resources [4]. Given the video represented by a set of terms, the similarity score is calculated using a TF-IDF algorithm.

The segment-level similarity uses Apache Lucene 3.6.2[3] to determine the relevance between the query anchor and the potential target segments. The Lucene standard analyzer was used with the default stop word list[4] to index ASR transcripts and manual subtitles. The search input query contained all the spoken data contained in the query anchor. The score calculation mechanism uses a combination of a Boolean AND function filter and ranking using the Vector Space Model [3]. The final score used to rank the hyperlinks was calculated by merging the two results as shown in Equation 1 and Equation 2.

$$Score = w_1 S_v + w_2 S_l \quad (1)$$

$$Score = S_v S_l \quad (2)$$

where S_v is the video-level similarity score, while S_l is the segment-level similarity score. We use a simple linear fusion mechanism to merge the two scores, where the weights w_1 and w_2 are set to 0.5 respectively.

2.3 Hyperlinking using Named Entities

This strategy links named entities contained in query anchors and the potential target segments, and then merges these entities to construct hyperlinks. Apache OpenNLP[5]

[1]https://github.com/dbpedia-spotlight
[2]http://dbpedia.org/Ontology
[3]http://lucene.apache.org/
[4]https://lucene.apache.org/core/3_6_2/api/core/org/apache/lucene/analysis/StopAnalyzer.html
[5]http://opennlp.apache.org/

Topic (Anchor) ID	4	12	21	23	27	31	32	39	43	45
MAP	0.8921	0.3733	0.1395	0.4925	0.0060	0.4170	0.5713	0.4127	0.1891	0.5555
P@5	1.000	1.000	0.600	1.0000	0.0000	0.8000	0.8000	1.0000	0.6000	1.0000
P@10	1.000	0.900	0.600	0.9000	0.0000	0.8000	0.7000	1.0000	0.7000	1.0000
P@20	1.000	0.700	0.550	0.8500	0.0000	0.6000	0.6000	0.8500	0.4000	0.9000

Table 1: Mean Average Precision (MAP) and P@N results for different topics in RUN 3

Run ID	Method	Data	Fuse
1	Text Annotation	M+I+S	Eq.1
2	Text Annotation	M+I+S	Eq.2
3	Text Annotation	M+U+S	Eq.1
4	Named Entities Link	M+L+S	Eq.4

Table 2: Run Details (M: Metadata, I: LIMSI, U: LIUM, S: Subtitle, Eq: Equation)

Run ID	1	2	3	4
MAP value	0.2944	0.2935	0.3109	0.0161
P@5	0.7000	0.7067	0.7267	0.0600
P@10	0.6567	0.6633	0.6567	0.1067
P@20	0.5450	0.5383	0.5433	0.0733

Table 3: Mean Average Precision (MAP) evaluation results

was used to tag words in the ASR transcripts and subtitles. All noun words tagged as *NN*, *NP*, and *NNP* were selected as named entities. To describe and link the named entities, a vector space model was constructed by predicting the surrounding words given the current word. We use word2vec[6] to implement a supervised learning mechanism using a Neural Net Language Model to create the vector model of named entities. We use the ASR transcripts of videos gathered from the blip10000 collection [7] as training data. The word2vec receives each named entity as input and outputs a vector $V = \{w_1, w_2, ...w_k\}$ where w_i is a surrounding word of the current entity learned by training data and the vector dimensionality k is set to 50, based on the experiment described in [5]. Equation 3 is used to calculate the score between different word vectors.

$$S = \frac{2(V_i \bigcap V_j)}{|V_i| + |V_j|} \tag{3}$$

where $V_i \bigcap V_j$ are the total number of words contained in both V_i and V_j. $|V_i|$ is the length of the word vector i. All named entities located at the potential target segments are merged using Equation 4 to generate the final score to obtain the ranked hyperlink list.

$$Score = \frac{\sum_{0<i<k} S_i}{N} \tag{4}$$

where S_i is the score of an entity in a potential target segment, and N is the total number of named entities in the current segment.

3. EXPERIMENTAL RESULTS

A total of four formal runs were submitted to the Search and Hyperlinking task in MediaEval 2013, described in Table 2. Table 3 shows the Mean Average Precision (MAP) value of each run. This indicates that our hyperlinking strategy based on spoken data annotation performs better. Table 1 shows P@N and MAP value of Run 3. MAP and P@N benchmark have received a good result in most runs except Topic (Anchor) 27, which describes Shakespeare and Global Theatre. A total of two other videos are related to Shakespeare and Global Theatre, while the content is presented in terms of a cartoon. The lack of visual elements leads to hyperlinks to cartoon segments, while real users will notice the unrelatedness between TV shows and cartoons.

[6]https://code.google.com/p/word2vec/

4. CONCLUSIONS

This paper presented details of DCU's participation in the TV Data Hyperlinking task of MediaEval 2013. The evaluation shows that annotating spoken data to construct hyperlinks achieves better results. In our future work, we will examine the use of visual cues to improve hyperlinking performance.

5. ACKNOWLEDGEMENT

This work is funded by the European Commission's Seventh Framework Programme (FP7) as part of the AXES project (ICT-269980).

6. REFERENCES

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

[2] L. Lamel and J.-L. Gauvain. Speech Processing for Audio Indexing. In *Advances in Natural Language Processing (LNCS 5221)*, pages 4–15. 2008.

[3] Lucene 3.6.2 Document. Apache Lucene - Scoring. https://lucene.apache.org/core/3_6_2/scoring.html, Dec. 2012.

[4] P. N. Mendes, M. Jakob, A. García-Silva, and C. Bizer. Dbpedia Spotlight: Shedding Light on the Web of Documents. In *Proceedings of the 7th International Conference on Semantic Systems*, USA, 2011.

[5] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient Estimation of Word Representations in Vector Space. In *Proceedings of Workshop at ICLR*, volume abs/1301.3781, 2013.

[6] A. Rousseau, F. Bougares, P. Dellsglise, H. Schwenk, and Y. Estssv. LIUM's systems for the IWSLT 2011 Speech Translation Tasks. In *Proceedings of IWSLT 2011I*, 2011.

[7] S. Schmiedeke, P. Xu, I. Ferrané, M. Eskevich, C. Kofler, M. A. Larson, Y. Estève, L. Lamel, G. J. F. Jones, and T. Sikora. Blip10000: A Social Video Dataset Containing SPUG Content for Tagging and Retrieval. In *Multimedia Systems Conference 2013, (MMSys '13)*, pages 96–101, 2013.