IDIAP RESEARCH REPORT

TOPIC-LEVEL EXTRACTIVE SUMMARIZATION OF LECTURES AND MEETINGS USING A SNIPPET SIMILARITY GRAPH

Chidansh A. Bhatt Andrei Popescu-Belis

Idiap-RR-09-2014

JUNE 2014

Centre du Parc, Rue Marconi 19, P.O. Box 592, CH - 1920 Martigny
T +41 27 721 77 11 F +41 27 721 77 12 info@idiap.ch www.idiap.ch

Topic-Level Extractive Summarization of Lectures and Meetings Using a Snippet Similarity Graph

Chidansh Bhatt and Andrei Popescu-Belis

June 13, 2014

Abstract

In this paper, we present an approach for topic-level video snippet-based extractive summarization, which relies on con tent-based recommendation techniques. We identify topic-level snippets using transcripts of all videos in the dataset and indexed these snippets globally in a word vector space. Generate snippet cosine similarity scores matrix, which are then utilized to compute top snippets to be utilized for summarization. We also compare the snippet similarity globally across all video snippets and locally within a video snippets. This approach has performed well on the AMI meeting corpus, in terms of ROUGE scores compare to state-of-the-art methods. Experiments showed that corpus like AMI meeting has large overlap between global and local snippet similarity of 80% and the ROUGE scores are comparable. Moreover, we applied proposed TopS summarizer in different scenarios on Video Lectures, to emphasize the merits of ease in utilizing summarizer with such content-based recommendation technique.

1 Introduction

State-of-the-art techniques for accessing video lectures (e.g., VideoLectures.NET, YouTube.com/edu or KhanAcademy.org), are mainly designed to facilitate browsing of a video and generate recommendations. Large collection of lectures and so does their recommendations, may lead to information overload and user would prefer to obtain maximal information about videos in shorter time. In our vision, a good audio-video summary will help the user obtain maximal information from the video, without having to watch the video lecture from beginning to the end. In recent years, video summarization has become an emerging field of research and we can categorize video summarization techniques in major categories like visual feature-based (e.g., motion, color, gestures, concepts), audio feature-based(e.g., speech transcript-based, audio-event), playable audio-visual materials or narratives or key frame-based story-board presentations etc.

In this work, our focus is on playable audio-visual summaries using speech transcripts of video lectures, emphasizing on topics discussed in the lecture. Also, we avoid the problems with coherence

that may arise with selection of single sentences in summary, rather we select the full topic-level segment/snippet to provide adequate context. Similar to lecture videos, we also consider the meeting videos for experiment and evaluation, as both have topic-level discourse.

2 Related Work

Summarization has been important and challenging research topics and it has large number of interesting works done. We can categorize them based on the different modalities like text, audio, video or combination of multiple modalities. There exist interesting survey based on each modalities (e.g., text summarization [15], video summarization [14] etc.). In this paper, we are focusing mainly on audio-visual summaries using speech transcripts. Thus, we provide a brief survey of closely related approaches.

Several methods have been proposed for ranking sentences based on some relevance metric. In supervised approaches [11], a classifier is trained on sentence features such as key words, sentence length, and position, and given a new text, a relevance score for each sentence is calculated. Also, unsupervised approaches like Maximal Marginal Relevance (MMR) [9, 18] is popular, where a term vector is created for each sentence. Other graph based methods include eigenvector centrality approaches that have been applied in TextRank [13] and LexRank [7] or an extension to such approaches like ClusterRank algorithm, where clustering or text-segmentation is to segment the transcript such that each section has utterances about the same subject. Our approach is more similar to ClusterRank approach and we emphasize the weaknesses of such approach in terms of issue with summary coherence, sentence redundancy and their performance evaluations.

3 Methodology

3.1 System Overview

The proposed TopS summarizer is represented in Figure 2. We generate topic-based segments/snippets from the ASR transcripts provided for each video lecture / AMI meeting. The topic segmentation was performed over the words using the TextTiling algorithm implemented in NLTK. We compute word-based similarity matrix between all snippets in the collection, using a vector space model and tf-idf weighting. We either consider the global (across the collection) or the local (within a video) similarity of each snippet to identify top most N relevant snippets to be incorporated in summary.

3.2 Topic-level Snippet Generation

Topic segmentation was performed over transcripts using TextTiling [10] as implemented in the NLTK toolkit [4] (available at http://nltk.org/). Topic shifts are determined based on the analysis of lexical co-occurrence patterns, which are computed from 20-word pseudo-sentences, to ensure uniform length.

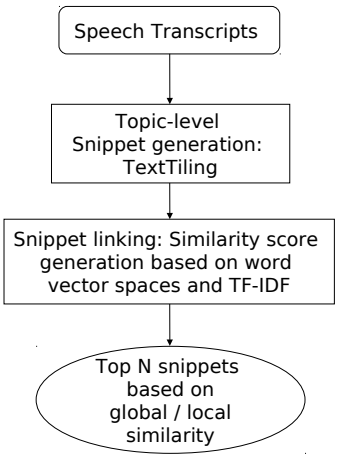

Figure 1: Overview of the proposed TopS Summarizer. Rectangles with sharp corners represent components of the system, while rounded boxes represent data.

Then, similarity scores are assigned at sentence gaps using block comparison. The peak differences between the scores are marked as boundaries, which we fit to the closest speech segment break. We selected TextTiling for its robustness and simplicity, although more advanced techniques such as TopicTiling [16] (same core algorithm but with LDA topic modeling) are also available.

3.3 Snippet Similarity Graph

We followed standard pre-processing to create the word vector representation, namely conversion to lower-case, tokenization, and stop word removal, using the NLTK library. We considered unigrams with size of the vocabulary 20k words. Each topic-level snippets' text segment is indexed in a word vector space with tf-idf weight [17]. In other words, the tf-idf weights w_{ij} for a given snippet i were (classically) computed as $w_{ij} = tf_{ij} \cdot idf_j$, where tf_j is the term frequency of word j in document d_i and idf_j is the inverse document frequency of word j. The similarity between two snippets $\vec{s}_i$ and $\vec{s}_j$ was then computed by the cosine similarity between them as follows:

$$sim_{cos}(\vec{s}_i, \vec{s}_j) = \frac{\vec{s}_i \cdot \vec{s}_j}{||\vec{s}_i||_2 \times ||\vec{s}_j||_2} \tag{1}$$

We generate M*M snippet similarity score matrix, where each row represents the snippet and columns represents their similarity scores to M snippets in the collection. Such similarity matrix can be used for generating content-based recommendation directly at the snippet level as well as the lecture level as shown in [1, 2, 3]. We process such similarity matrix to generate the snippet similarity graph, where each node of the graph is the snippet and each link is representing its similarity score to the connecting snippet. More precisely, a directed edge from snippet X to snippet Y is weighted by their cosine similarity score retrieved from the similarity matrix. An important difference here compared to other existing graph

based approaches where the sentence level graphs are generated, proposed graph is at snippet level. This will help avoiding coherence and sentence redundancy issues by providing snippet level summary rather than sentence level summary. Even such approach will not require further redundancy check at sentence level.

Also, while many of the graph based approaches ignores or removes the edges with zero value, we utilize them for further computation. All the snippets in the collection is utilized to determine the most important snippets within a video to be linked to generate the summary. We also map snippet ids to corresponding video ids for reference.

3.4 Snippet-based Summarization

Once snippet similarity graph is generated, we considered two scenarios to compute the importance of the node. Main motivation behind consideration of two different scenarios arise from the fact that, Klewel has public and private customers and in turn their corpus is large collection of video lectures within a conference or across conferences on similar topics or sometimes isolated lectures on diverse topics for some private customers. In the first scenario, called global similarity based linking/summarization, we assume that snippets that are related to many other snippets in the collection are likely to be central and have high weight for selection in summary. For such global similarity based linking, all the snippets in the collection contribute to the total similarity score of a snippet, as we compute the relevance_measure for each snippet S as the summation of its similarity scores over all the M nodes in the complete graph,

$$Relevance\_measure(S) = \sum_{i=1}^{M} sim\_score(S, i) \tag{2}$$

Each snippet‘s relevance_measure is considered while linking them in the summary, assuming that snippets that are related to many other snippets with good similarity score in the collection are likely to have larger relevance_measure compared to the snippets with a few or low similarity scores in the complete graph. Though in the second scenario, called local similarity based linking, we emphasize on possibility of diverse topics that are not much frequent in the entire collection and thus required to be computed within subset of the entire collection or even at individual level. In particular, we compute the relevance_measure for each snippet S_l of video l as the summation of its similarity scores over all the K nodes mapped as snippets of video l in the graph,

$$Relevance\_measure(S_l) = \sum_{i=1}^{K} sim\_score(S_l, i) \tag{3}$$

Even in the hybrid approach such local and global similarities can be given different weighting to come up with more suitable relevance measure score for snippets. We will experiment such hybrid approach in future. For now, once summation is done, all the snippets within video are sorted by their relevance measure scores. Based on user selected threshold P% for summary size (e.g., 10%, 17%, etc.) respect to total number of snippets K in the video we calculate the value $N = (100)*(P/K)$. As indicated in [12],

summaries as short as 17% of the full text length speed up decision making twice, with no significant degradation in accuracy. Once top N snippets are selected, they will be linked by their actual temporal order in the video to generate the summary. For the smooth transition between the linked top N snippets we inserted the fading effects in between every link for the summary video.

It was also interesting to observe if the snippets with lowest similarity scores can be more informative compared to the snippets with highest similarity scores. Thus, we also did experiments with selecting top N snippets with lowest similarity scores and linked by temporal order to generate the summary video. In future we plan to also consider edges with multiple weights (e.g., similarity based on random projection or latent semantic indexing together with TF*IDF). Though, in our earlier works we observed that TF*IDF based similarity worked reasonable for snippet level content-based recommendation [2].

TopS provides freedom to consider the snippet similarity globally or locally. We submit that global similarity choices can be utilized in summarization of video lectures within a conference or across conferences on similar topics, while local similarity is more suitable for isolated lectures on diverse topics.

4 Experimental Results

Proposed TopS summarizer has been evaluated on the AMI meeting corpus [6] and Klewel Video Lectures. The dataset and evaluation metrics are described in following sections. We provide below our results and a detailed discussion on achieved performance compared to other observed results, as well as our additional testing to provide view on the ease of using proposed TopS summarizer.

4.1 Dataset

In this work, we use AMI meeting data sets that have manually annotated summaries and for the Klewel Video Lectures we do not have manually annotated summaries.

The AMI meeting corpus is a collection of 100 hours of meeting data that includes speech audio, transcripts, and human summaries. Each meeting has participants talking for about 35 minutes on a given topic and the transcripts are about 3000- 7000 words. Our experiments are based on the 137 meetings that have extractive human summaries.

Klewel video lectures is a collection of total 251 talk from 8 events (e.g., lectures at conferences, project meetings, class lectures, etc.), out of which 169 videos have ASRs available. Though, it do not have human generated extractive summaries.

4.2 Evaluation Metrics

ROUGE is used to evaluate the performance for all the systems. ROUGE scores are based on the number of overlapping units such as n-grams, between the system generated summary and the ideal summaries created by humans.

4.3 Results and Discussion

For each meeting, we generate an extractive summary satisfying a length constraint specified in terms of a percentage (e.g., 17%) of the total number of snippets within a video lecture to be summarized. We consider three different scenarios for experiments on AMI meeting corpus. (1) TopS with global similarity based on highest scored links: we consider the highest summation scores to qualify the snippet for including in the summary. (2) TopS with global similarity based on lowest scored links: we consider the lowest summation scores to qualify the snippet for inclusion in the summary and (3) TopS with local similarity based on highest scored links: we consider the highest summation scores only among the snippets within a video to qualify the snippet for including in the summary. Also, other state-of-the-art methods like TextRank [13] and ClusterRank [8].

Global similarity based on highest scored links											
ROUGE 1			**ROUGE 2**			**ROUGE 3**			**ROUGE 4**		
Pre.	Rec.	F	Pre.	Rec.	F	Pre.	Rec.	F	Pre.	Rec.	F
0.70	**0.59**	**0.61**	0.42	**0.34**	**0.36**	**0.30**	**0.24**	**0.26**	0.26	**0.21**	**0.22**
Global similarity based on lowest scored links											
ROUGE 1			**ROUGE 2**			**ROUGE 3**			**ROUGE 4**		
Pre.	Rec.	F	Pre.	Rec.	F	Pre.	Rec.	F	Pre.	Rec.	F
0.81	0.12	0.21	**0.43**	0.06	0.11	0.27	0.03	0.06	0.23	0.03	0.05
Local similarity based on highest scored links											
ROUGE 1			**ROUGE 2**			**ROUGE 3**			**ROUGE 4**		
Pre.	Rec.	F	Pre.	Rec.	F	Pre.	Rec.	F	Pre.	Rec.	F
0.70	**0.59**	**0.61**	0.42	**0.34**	**0.36**	**0.30**	**0.24**	**0.26**	**0.27**	**0.21**	**0.22**
ClusterRank with normalization											
ROUGE 1			**ROUGE 2**			**ROUGE 3**			**ROUGE 4**		
Pre.	Rec.	F	Pre.	Rec.	F	Pre.	Rec.	F	Pre.	Rec.	F
0.32	0.26	0.28	0.05	0.04	0.04	-	-	-	-	-	-
TextRank with cosine similarity											
ROUGE 1			**ROUGE 2**			**ROUGE 3**			**ROUGE 4**		
Pre.	Rec.	F	Pre.	Rec.	F	Pre.	Rec.	F	Pre.	Rec.	F
0.30	0.21	0.24	0.05	0.03	0.04	-	-	-	-	-	-

Table 1: Result with precision , recall and f-measure of ROUGE for TopS, ClusterRank and TextRank methods on AMI meeting corpus.Scores in the bold represent the highest among all the methods

As shown in the Table 1, the TopS summarizer outperformed ClusterRannk [8] and TextRank [13] based on graph model and PageRank [5] approaches with 110% improvement in terms of average ROUGE score. It is interesting to observe that TopS with consideration of lowest similarity scores based linking has higher precision for ROUGE-1 and ROUGE-2, but gradually its performances drastically reduces. On the other hand, TopS with highest similarity score based linking has consistently better performance. Also, in case of AMI meeting corpus it is very supportive to observe that local and global similarity scores based TopS is performing similarly due to the topic similarity across AMI meeting corpus. AMI meetings are captured under same topic and thus consideration of global or local similarity do not deviate much the highly relevant snippet, we onserved that there are 80% overlap in the snippets linked in global and local scenarios.

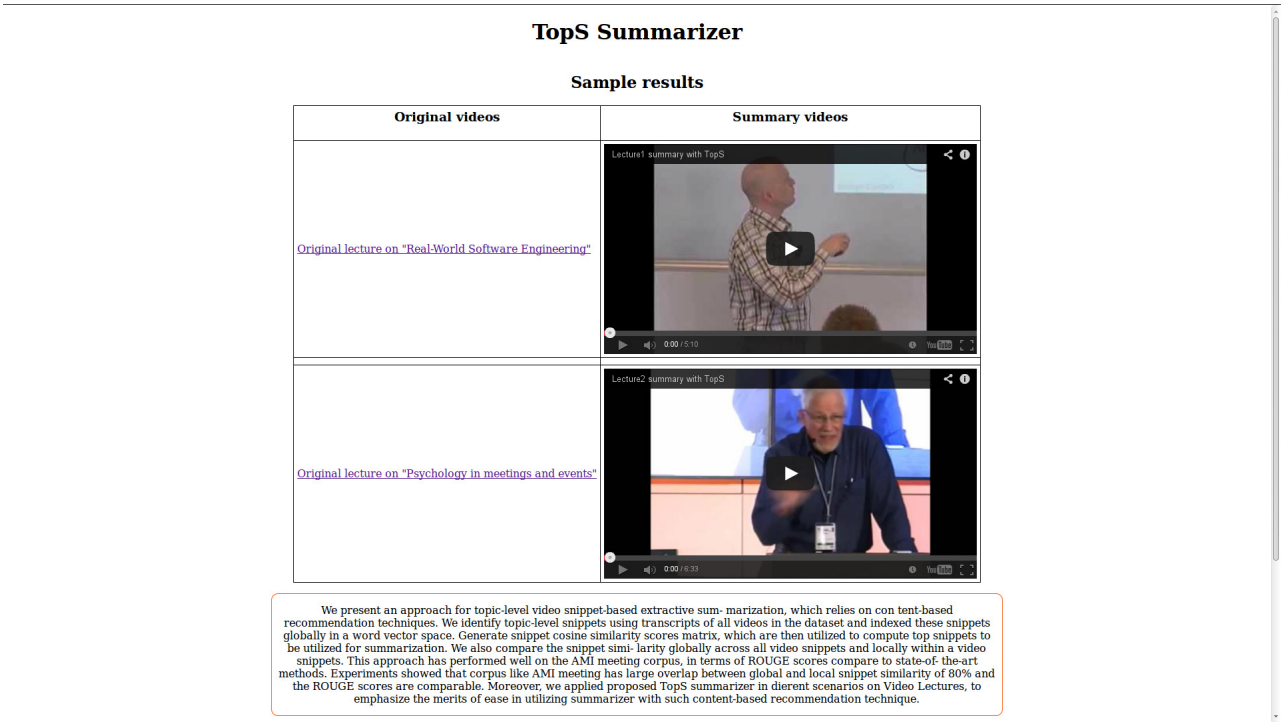

Figure 2: Screen shot of TopS Summarizer

Difference in global and local similarity based TopS will be more evident and interesting for the klewel dataset due to diversity in their topics. We have provided a demonstrator [1] of the results of the TopS summarizer on several talks available from klewel dataset.

5 Conclusion and Perspectives

We have presented a topic-level graph-based extractive summarization of lectures, which provides a solution for providing easy and fast access to information within each video using topic-level segments

[1] http://www.idiap.ch/c̆bhatt/Demo_TopS_summarizer.html

of transcripts. TopS leverage on topic-model, content-based recommendation as well as graph method to provide novel solution avoiding the problems with coherence and sentence redundancy that may arise with selection of single sentences in summary. Provided freedom to consider the snippet similarity globally or locally has enable potential applications to large and diverse set of datasets with similar or different topics. Overall simplicity in approach, ease in integration with existing recommendation systems and better performance compared to state of the art methods makes TopS a much interesting summarization approach.

6 Acknowledgments

This work was supported by the Swiss National Science Foundation (AROLES project n. 51NF40-144627).

References

[1] C. Bhatt, A. Popescu-Belis, M. Habibi, S. Ingram, F. McInnes, S. Masneri, N. Pappas, and O. Schreer. Multi-factor segmentation for topic visualization and recommendation: the MUST-VIS system. In *Proceedings of ACM Multimedia 2013, Grand Challenge Solutions*, pages 37–42, Barcelona, 2013.

[2] Chidansh Bhatt, Nikolaos Pappas, Maryam Habibi, and Andrei Popescu-Belis. Idiap at MediaEval 2013: Search and hyperlinking task. In Martha A. Larson, Xavier Anguera, Timo Reuter, Gareth J. F. Jones, Bogdan Ionescu, Markus Schedl, Tomas Piatrik, Claudia Hauff, and Mohammad Soleymani, editors, *MediaEval*, volume 1043 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2013.

[3] Chidansh Bhatt, Nikolaos Pappas, Maryam Habibi, and Andrei Popescu-Belis. Multimodal reranking of content-based recommendations for hyperlinking video snippets. In *Proceedings of International Conference on Multimedia Retrieval*, ICMR '14, pages 225:225–225:232, New York, NY, USA, 2014. ACM.

[4] Steven Bird. NLTK: the Natural Language Toolkit. In *COLING/ACL Interactive Presentations*, Sydney, 2006.

[5] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.*, 30(1-7):107–117, April 1998.

[6] Jean Carletta, Simone Ashby, Sebastien Bourban, Mike Flynn, Mael Guillemot, Thomas Hain, Jaroslav Kadlec, Vasilis Karaiskos, Wessel Kraaij, Melissa Kronenthal, Guillaume Lathoud, Mike Lincoln, Agnes Lisowska, Iain McCowan, Wilfried Post, Dennis Reidsma, and Pierre Wellner. The ami meeting corpus: A pre-announcement. In *Proceedings of the Second International Conference on Machine Learning for Multimodal Interaction*, MLMI'05, pages 28–39, Berlin, Heidelberg, 2006. Springer-Verlag.

[7] Günes Erkan and Dragomir R Radev. Lexrank: Graph-based lexical centrality as salience in text summarization. 2004.

[8] Nikhil Garg, Benot Favre, Korbinian Riedhammer, and Dilek Hakkani-Tr. Clusterrank: a graph based method for meeting summarization. In *INTERSPEECH*, pages 1499–1502. ISCA, 2009.

[9] Jade Goldstein, Mark Kantrowitz, Vibhu Mittal, and Jaime Carbonell. Summarizing text documents: Sentence selection and evaluation metrics. In *Proceedings of the 22Nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '99, pages 121–128, New York, NY, USA, 1999. ACM.

[10] Marti A. Hearst. TextTiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics*, 23(1):33–64, 1997.

[11] Julian Kupiec, Jan Pedersen, and Francine Chen. A trainable document summarizer. In *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '95, pages 68–73, New York, NY, USA, 1995. ACM.

[12] Inderjeet Mani, Gary Klein, David House, Lynette Hirschman, Therese Firmin, and Beth Sundheim. Summac: A text summarization evaluation. *Nat. Lang. Eng.*, 8(1):43–68, March 2002.

[13] Rada Mihalcea and Paul Tarau. Textrank: Bringing order into texts. In Dekang Lin and Dekai Wu, editors, *Proceedings of EMNLP 2004*, pages 404–411, Barcelona, Spain, July 2004. Association for Computational Linguistics.

[14] Arthur G. Money and Harry Agius. Video summarisation: A conceptual framework and survey of the state of the art. *J. Vis. Comun. Image Represent.*, 19(2):121–143, February 2008.

[15] Ani Nenkova and Kathleen McKeown. A survey of text summarization techniques, 2012.

[16] Martin Riedl and Chris Biemann. TopicTiling: a text segmentation algorithm based on LDA. In *Proceedings of the ACL 2012 Student Research Workshop*, pages 37–42. Association for Computational Linguistics, 2012.

[17] Gerard Salton and Christopher Buckley. Term-weighting approaches in automatic text retrieval. *Information Processing and Management Journal*, 24(5):513–523, 1988.

[18] Dingding Wang, Shenghuo Zhu, Tao Li, and Yihong Gong. Comparative document summarization via discriminative sentence selection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 7(1):2, 2013.