Leveraging Multimodal LDA for Hyperlinking

Anca Roxana Simon, Ronan Sicre, Rémi Bois, Guillaume Gravier, Pascale Sébillot

IRISA

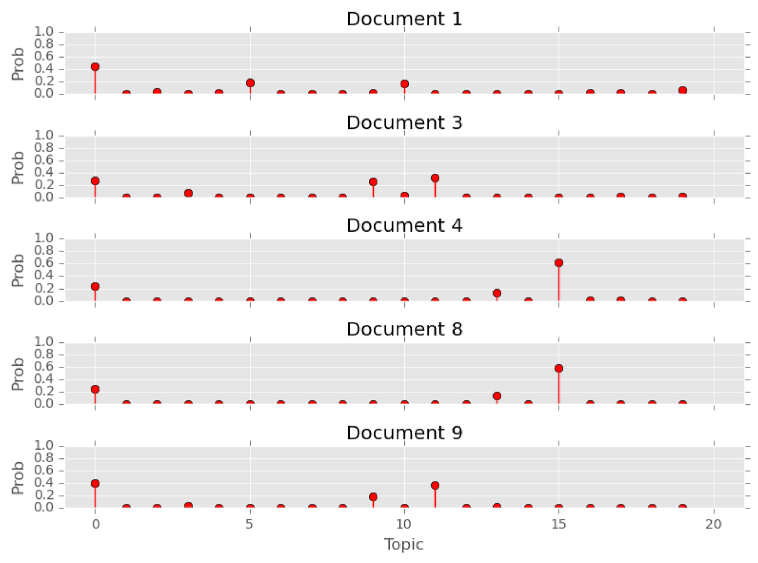
Abstract

We present the runs that we submitted to the Video Hyperlinking task. Two out of four runs use cross-modal Latent Dirichlet Allocation (LDA) as a means to jointly use visual and audio information in the videos. As a contrast, one run is based solely on visual information, and the last one is a combination of cross-modal and visual runs.

LDA for hyperlinking

LDA learns latent topics from a series of documents. Documents are associated with a distribution of topics, used to link them.

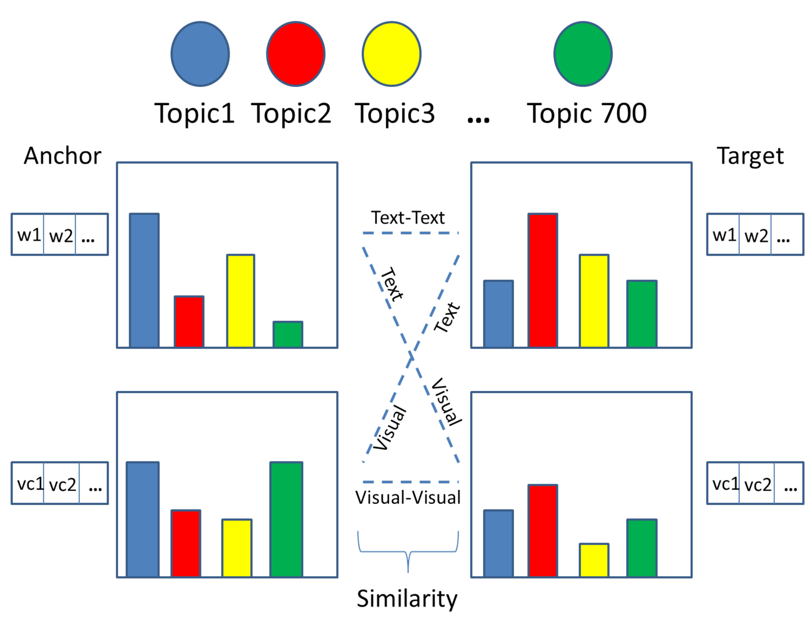
LDA example



- High similarity scores can be achieved by documents that do not share much vocabulary

Building a Bimodal LDA

- Originally used with two languages in machine translation [2]
- Adapted to two modalities (audio and visual informations)
- 700 latent bimodal topics were extracted
- Objectives : diversity and serendipity [1]
- Four different kinds of links built depending on the modalities



Most frequent words in both modalities for two topics

Topic 3 Audio love home feel day life baby made thing la

Topic 3 Visual singer microphone sax concert master-of-ceremonies cornet flute

Topic 25 Audio years technology computer find key future power machine speed

Topic 25 Visual equipment machine tape-player computer appliance-recording

Experiments

Data and Runs

- Closest to a real-word setting
- LIMSI's automatic transcriptions are used
- Leuven's extracted visual concepts are used

Run1 Visual concepts similarity (no topics) with visual reranking (on top 50)

Run2 Audio to visual with visual reranking (on top 50)

Run3 Visual to audio with ngram reranking (on top 50)

Run4 Rank Aggregation (with Runs 1-2-3 and a pure ngram scoring)

Results

	Minimum	25%	50%	75%	Maximum
Prec 10	0.017	0.198	0.275	0.524	0.608
Run1		0.207			
Run2	0.017				
Run3		0.224			
Run4	0.156				

We attribute the low scores to our real-world setting choice as well as some properties of the evaluation (see thereafter).

Discussion – Evaluation

The evaluation has been caried via Mechanical Turk, a realistic and affordable way to obtain a large amount of relevance score. However, we notice a few cons:

- Only one evaluation per anchor-target pair
- Yes/No question on relevance
- No clue on the difficulty of the task

We evaluated the Fleiss-κ on a similar task and obtained very low scores (near 0), indicating that anchor-target relevance is highly subjective.

Discussion – Diversity and serendipity

Despite forbidding the use of the same show for target candidates, many near-duplicates remain due to rebroadcasted shows. These near-duplicates are not interesting for users who are mostly looking for new information.

Conclusion

We proposed a new way to link video fragments. This new method focuses on diversity and serendipity, two aspects that are not evaluated in the hyperlinking task. Some low results can be attributed to the preference for highly similar anchor-target pairs in the evaluation.

References

[1] Ilaria Bordino, Yelena Mejova, and Mounia Lalmas. Penguins in sweaters, or serendipitous entity search on user-generated content. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 109–118. ACM, 2013.

[2] Wim De Smet and Marie-Francine Moens. Cross-language linking of news stories on the web using interlingual topic modelling. In *2nd ACM Workshop on Social Web Search and Mining*, 2009.