ComInLabs

Hierarchical Topic Models
for
Language-based Video Hyperlinking

Anca-Roxana Şimon, Rémi Bois, Guillaume Gravier
Pascale Sébillot, Emmanuel Morin, Sien Moens

UMR IRISA

lina

DEPARTEMENT COMPUTER WETENSCHAPPEN

cnrs

UNIVERSITÉ DE RENNES 1

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES RENNES

informatiques mathématiques

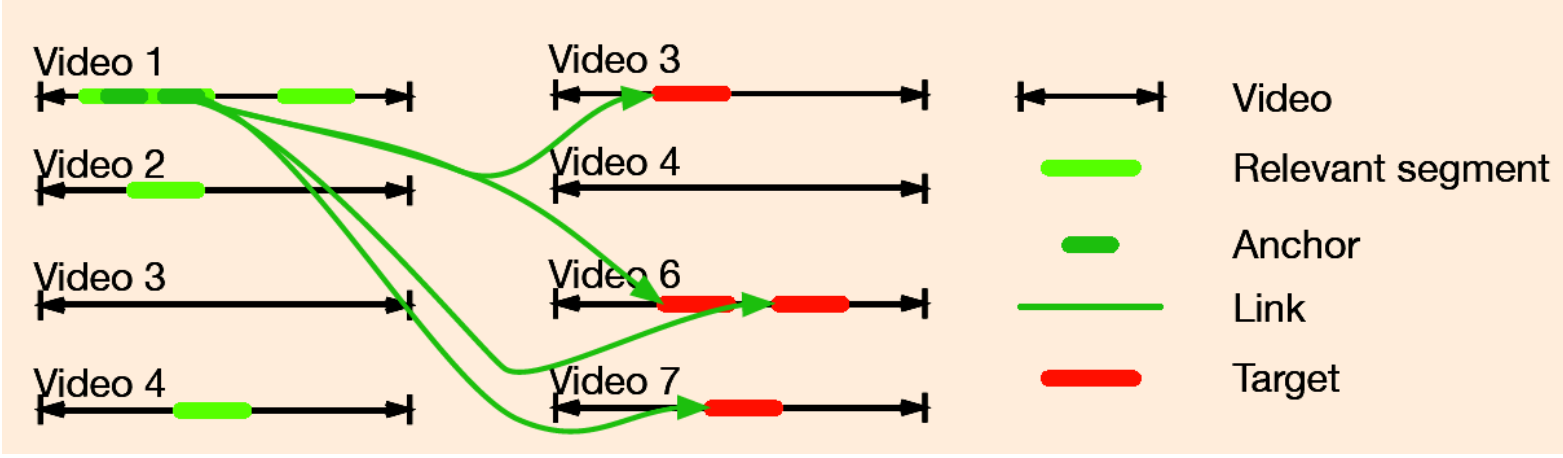
UNIVERSITÉ DE NANTES

KU LEUVEN

Video hyperlinking

The (search and) hyperlinking task scenario:

1. search: answering a query with a ranked list of documents
2. anchor detection: finding potential anchors in the videos
3. linking: creating a ranked list of segments related to the anchors



Implemented within Mediaeval from 2012 to 2014, TRECVid task since 2015

Beyond search,

anchor detection + hyperlinking = organizing a collection

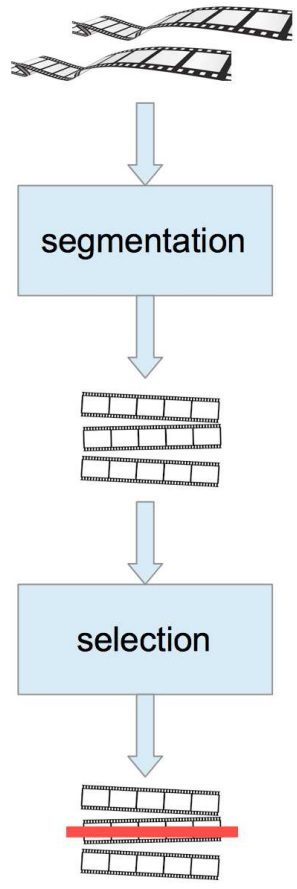
for analytics based on interaction with the data

M. Eskevich *et al.* Multimedia information seeking through search and hyperlinking. ICMR, 2013

An overview of the state of the art

1. **Segmentation**
   - fixed-length segments
   - video shots
   - topic segments
   - utterance
2. **Target selection**
   - language via transcripts

     $\longrightarrow$ possibly enriched with entities, prosody, etc.
   - visual content (often via concepts)
   - metadata



mostly direct comparison in vector space with cosine similarity!

The need for diversity (and serendipity)

Direct content comparison = targets very similar to the anchor

- high-similarity = low-risk evaluation (because of the lack of (known) goal)
- ngram ($n \in [1, 4]$) alignment worked very well in 2014

Safe targets obtained from direct content comparison are

- near duplicates — limited interest
- timeline events — that's better

but lacks **diversity to cover potential users interests** and enable serendipity.

Use hierarchical topic models to control diversity

- can link anchor/target pairs with few words in common
- can control and increase diversity
- can be used to explain the nature of the link (not in the paper)

Outline

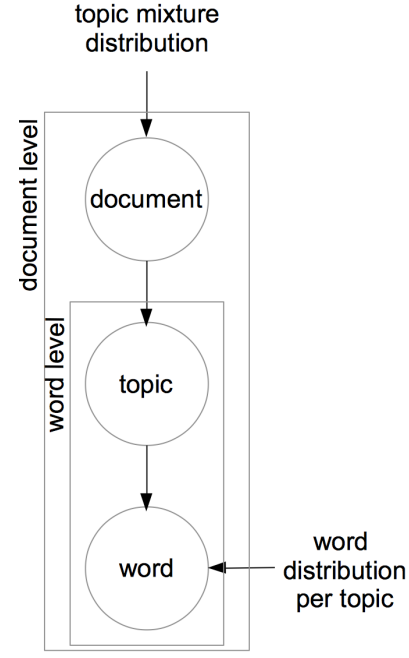
- Hierarchical topic models
  - Independent topic levels
  - Tree-structured topic levels
- Experimental evaluation
  - Data and metrics
  - Results
- Discussion

LDA topic models

Principle: Explain documents in a collection as a mixture of K topics, where each word is assigned to a topic.

I eat fish and vegetables.
Fishes are pets.
My kitten eats fish.
[source: wikipedia]

The K topics are learned from the data and are each characterized by a probability distribution z_i over the vocabulary



Hierarchy with 10 levels, trained independently on the transcripts of the BBC video collection (details later)

$K \in \{50, 100, 150, 200, 300, 500, 700, 1000, 1500, 1700\}$

- level 1, $K_1 = 50$, broad topics $z_i^1 \quad (i \in [1, K_1])$
- level 10, $K_{10} = 1700$, fine-grain topics z_i^{10}

All words appear in all topics at every level, only with different probabilities

LDA topic models illustrated

LDA training with Gibbs sampling with standard values for the hyperparameters $\alpha = 50/K$ and $\beta = 0.01$.

$z_1^1 \quad z_2^1 \quad z_3^1 \quad \dots \quad z_{50}^1$

$z_1^2 \quad z_2^2 \quad z_3^2 \quad \dots \quad z_{50}^2 \quad \dots \quad z_{100}^2$

$\vdots$

$z_1^7 \quad z_2^7 \quad z_3^7 \quad \dots \quad z_{50}^7 \quad \dots \quad z_{100}^7 \quad \dots \quad z_{700}^7$

z_1^6	z_2^6		z_{500}^6
team	island		animal
football	sea		herd
player	coast		lion
game	place	...	elephant
sport	mile		hunting
cricket	water		africa
league	road		food
england	beach		calf

Independent levels

Use segment probabilities given by topic models at different levels

$\rightarrow$ independently one from another

$\rightarrow$ combined across levels

New representation of a segment x at level l (sparse version: 10 non null entries)

$$x_l = \left(p(x|z_1^l) \dots p(x|z_{K_l}^l)\right)$$

where

$$p(x|z_i^l) = \sqrt[n_x]{\prod_{j=1}^{n_x} p(w_j|z_i^l)} \quad \text{with} \quad p(w_j|z_i^l) = \frac{n(z_i^l, w_j) + \beta}{\sum_{k=1}^{n} n(z_i^l, w_k) + \beta|V|}$$

$n(z_i^l, w)$ = number of times w is associated with topic z_i^l

Independent levels (cont'd)

$$S_1(x, y) = \sum_l \alpha_l \log (x_l \cdot y_l)$$

IT_k	only level k	$\alpha_k = 1, \alpha_{i \neq k} = 0$
$\text{IT}_=$	equal weights	$\alpha_k = 0.2 \quad \forall k \in \{1, 3, 5, 7, 9\}$
$\text{IT}_<$	general < specific	$\alpha_1 = 0.1, \alpha_3 = 0.15, \alpha_5 = 0.2, \alpha_7 = 0.25, \alpha_9 = 0.3$
$\text{IT}_>$	specific < general	$\alpha_1 = 0.3, \alpha_3 = 0.25, \alpha_5 = 0.2, \alpha_7 = 0.15, \alpha_9 = 0.1$

Combination considered only over 5 levels (no use to consider more than 5 levels)

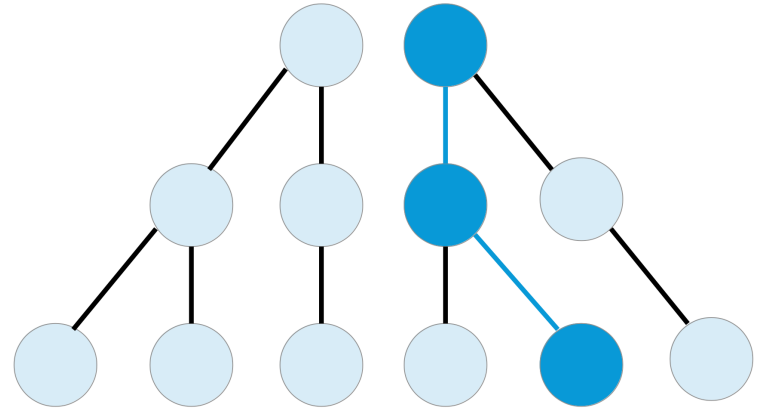
Tree-structured topics

Use path across a hierarchy of topics to caracterize a segment

$\rightarrow$ direct topic linking

$\rightarrow$ global topic linking

$$S_2(x, y) = \sum_l \alpha_l \log p(y|t_x^l)$$



with t_x^l the topic distribution at level l in the best path for x

x = anchor, y = target

Tree-structured topics (cont'd)

Direct linking

- link z_i^l to the closest topic at $l-1$, i.e., $z_k^{(l-1)}$ s.t. $k = \arg\max_j z_i^l \cdot z_j^{l-1}$
- not necessarily sibling or child

Global linking

- link similar topics at consecutive levels with contraints
  - every node has one parent
  - each node has at least two children
- solved between consecutive levels with ILP

$$\max \sum_{i=1}^{K_l} \sum_{j=1}^{K_{l+1}} z_i^l \cdot z_j^{l+1} \, \text{link}(i,j)$$

$$\text{subject to} \sum_{i \in [1,K_l]} \text{link}(i,j) = 1 \quad \text{and} \sum_{j \in [1,K_{l+1}]} \text{link}(i,j) \geq 2$$

Data and metrics

Mediaeval 2013 & 2014 Search and Hyperlinking data

- 4,000 h of BBC broadcast data, approx. 45 min each
- automatic speech transcripts (courtesy of LIMSI)
- 30 anchors in 2013, 30 anchors in 2014

Task considered: reranking targets

- targets are taken from the participants' submissions
- relevance judgments provided by turkers

year	anchor duration	#targets (% relevant)	target duration 95 % interval
2013	32.2	9,973 (29.9 %)	[82.58, 84.18]
2014	22.9	12,340 (15.3 %)	[58.12, 59.58]

Targets come from a variey of systems (textual content, visual content, or a combination of both possibly with additional sources (metadata, prosody, etc.)) but were *all proposed for a reason*!

Evaluating each level

	2013			2014		
mAP	@5	@10	@20	@5	@10	@20
DirectH	0.71	0.66	0.62	0.41	0.41	0.38
IT_{50}	0.63	0.62	0.59	0.39	0.36	0.33
IT_{150}	0.67	0.64	0.58	0.45	0.4	0.35
IT_{300}	0.64	0.6	0.58	0.33	0.34	0.32
IT_{700}	0.64	0.61	0.58	0.34	0.3	0.32
IT_{1500}	0.62	0.6	0.55	0.41	0.4	0.37

- no statistical difference between direct and topic (paired t-test, $p < 0.05$)
  $\rightarrow$ but improving is not the goal here
- no statistical difference between topic granularity
- anchors with relevant targets are addressing general topics

Evaluating each level from a different angle

Tolerance precision after 15 s of the top-10 targets seen as an entry point to relevant fragments (p_10_tol)

year	number of topics (K)				
	direct	50	150	300	700
2013	0.25	**0.44**	**0.34**	**0.35**	**0.34**
2014	0.19	0.18	**0.25**	**0.26**	0.21

- topic-based ranking provides better entry points ...
- ... but hard to define the appropriate number of topics

Combining levels

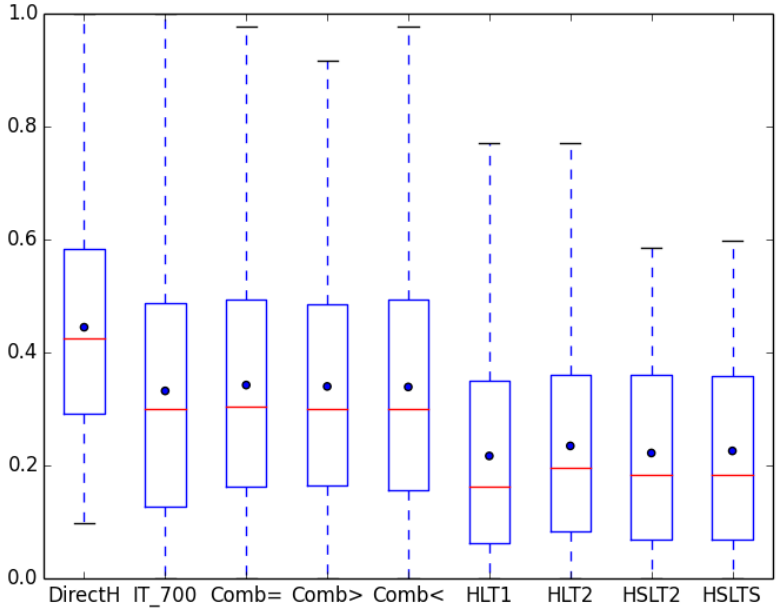
	2013			2014		
	@5	@10	@20	@5	@10	@20
DirectH	0.71	0.66	0.62	0.41	0.41	0.38
IT_{150}	0.67	0.64	0.58	0.45	0.4	0.35
Independent topic levels (=)	0.7	0.67	0.63	0.34	0.33	0.31
Independent topic levels (<)	0.68	0.66	0.62	0.31	0.33	0.32
Independent topic levels (>)	0.71	0.68	0.63	0.35	0.35	0.33
Direct topic links (5 levels)	0.54	0.49	0.43	0.43	0.38	0.35
Direct topic links (4 levels)	0.44	0.44	0.39	0.43	0.43	0.37
ILP topic links (4 levels)	0.4	0.39	0.37	0.43	0.44	0.41

- combining levels has no effect in 2013
- combining independent topics is detrimental in 2014
- using tree topic structures helps in 2014

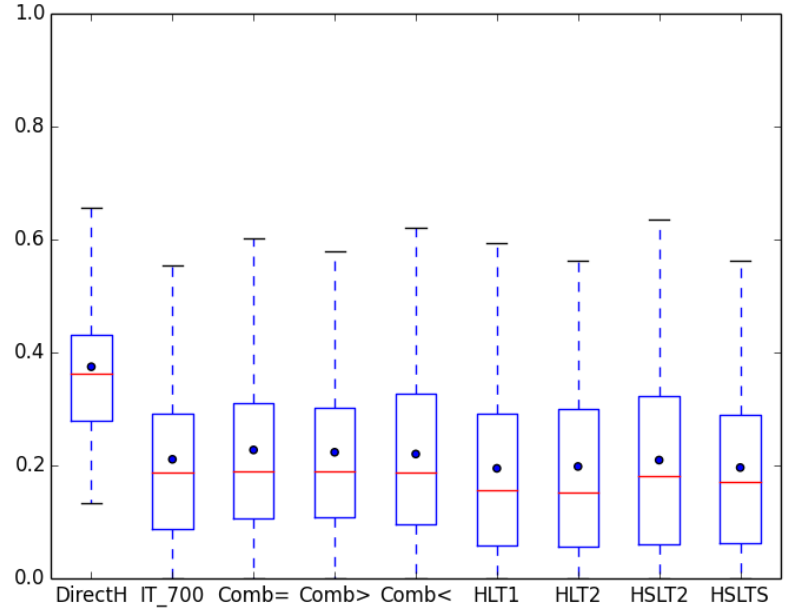
shorter and more realistic anchors in 2014 + absence of context?

Interest of topic hierarchies

Cosine distance between top-20 relevant anchor-target pairs for each system shows that we can capture vocabulary diversity (globally diversity also?)



2013 data



2014 data

Interest of topic hierarchies (cont'd)

A large proportion of the top-20 relevant links differ between systems!

System 1	System 2	% difference 2013	% difference 2014
700 LDA topics	direct cosine	93	86
700 LDA topics	independent levels (>)	82	90
700 LDA topics	direct topic linking	98	93
independent levels (=)	ILP topic linking	94	95

Interest of topic hierarchies: A case study

Anchor:

William the Conqueror. He parcelled out the country to the leading families who had fought for him. To control their enormous estates, they built the first stone castles in England. They were the power bases of the second order of society, the military aristocracy. The medieval world was studded with castles, hundreds of them. “The bones of the kingdom”, as one contemporary called them. They were built to be high, to act as giant watchtowers over the surrounding countryside. To see, and to be seen.

Direct linking with cosine:

A stone castle like this would be the biggest, most expensive and most threatening building you’d be likely to see in your life. It was a symbol of the power of the aristocracy, the centre of their great estates and the foundation of their military might.

Sibling topic linking:

I’m on my way to the site of the biggest castle in England. It must also rank as one of the very oddest in the whole of medieval Britain. It was built around 1313 by a colorful character called Thomas of Lancaster. In his day, Thomas was talked about even more than his cousin, who happened to be none other than the King of England, Edward II. Thomas fell out spectacularly with the king when he murdered one of Edward’s closest friends. It was then that Thomas built this Dunstanburgh Castle.

Interest of topic hierarchies: A case study

Best topic	Sibling topic	General topic
city	great	people
people	city	world
place	empire	war
good	roman	city
countryside	world	british
heart	christian	britain
centre	building	life
nation	living	great
visit	light	work
capital	modern	history

In summary

Topic models can help provide diversity in NLP-based video hyperlinking, potentially favoring serendipity, but many things remain to be done, including

- combining all this with visual concepts
  - $\rightarrow$ bimodal topic models at TRECVid video hyperlinking
- evaluation that accounts for diversity
  - $\rightarrow$ short study indicates that we have diversity
  - $\rightarrow$ can we extend to large-scale evals?
- applications leveraging the diversity (analytics but how?)
- explanation of the links based on topic relations
  - $\rightarrow$ should help link acceptability and serendipity