

Analyzing and mining multimedia collections

How (can) statistics help?

Guillaume Gravier



Multimedia data: what and what for?

Multimedia data = all kind of digital data related to media, i.e., intended for communication of a(n informative) message towards human

A wide variety of sources

→ TV, radio, newspapers, web media; photo agencies; video/photo sharing sites; social media; etc.



A wide variety of material

→ texts, posts and tweets; audio, speech and music; images; videos; etc.



A wide variety of application domains

→ media asset management; Internet portals; connected TV; on-line courses; etc.



Multimedia data are...

1. semantic

- created by humans for humans
- make sense only in context
- variability, heterogeneity and multimodality

⇒ *robust and versatile machine interpretation*



2. very very large scale

- 100 h of videos / min on YouTube!
- 58 billion tweets / day

⇒ *efficient processing and data organization*

3. unstructured and disconnected

- implicit structure not known ⇒ limited interpretation capabilities
- lack of organization and links within collections

⇒ *collection structuring methodology*

Multimedia content processing: the big data era

From content description...

- all-purpose descriptors for texts, images, sounds and videos
→ lemma/stem, MFCC, SIFT, VLAD, Fisher vector, ...
- supervised machine learning to detect (semantic) concepts
→ SVM, HMM, CRF, MKL, DNN, ...
- information retrieval to search and rank documents at scale

... to analytics

Data analytics can be defined as the application of [] data processing techniques to discover patterns, extract knowledge and gain insights from large-scale, typically multi-source data collections that may contain structured, unstructured and semi-structured data. — excerpt from BDVA Strategic Research and Innovation Agenda

Current limits and challenges

- semantic **interpretation still far from perfect**
 - multimodality, structured I/O, confidence, annotation, etc.
- multimedia **pattern mining at scale is lacking**
 - object discovery, audio pattern discovery, etc.
- multimedia **data collections are unstructured**
 - multimedia data warehouses, graph structures, etc.
- **usage and expectations are still unclear**
 - practical use of technology, acceptability, evaluation, etc.
- **privacy and security are challenged** by analytics and data agregation

Texmix: description (and browsing)

The screenshot displays the Texmix web interface, a navigation tool for broadcast news collections. The interface is organized into several sections:


- Header:** Includes navigation links (Home, Hypervideos, Dynamic Resume, Geotagging), a search bar, and the date 28/03/2007. A timeline at the top shows a sequence of dates from 27/03/07 to 31/03/07, with a 'TODAY' marker.
- Summary (2):** A row of small video thumbnails representing different news segments.
- Audio Resume (4):** A red waveform visualization of the audio content.
- Report 3 (7):** A section titled 'LAGUILLER ARLETTE SÉGOLÈNE IPSOS MARIE INSÉCURITÉ SARKOZY VERDIER CAMPAGNE'.
- See Also (5):** A list of related news items with dates and report numbers, such as '14/03/2007 report 2' and '17/03/2007 report 3'.
- Related Videos (5):** A list of related video reports with dates and report numbers.
- Video Player (1):** A video player showing a news segment with a progress bar and a play button. The video title is 'dix huit et demi jean marie le pen treize pour cent'.
- Disable Subtitles (2) and Similar Images (6):** Interactive buttons for video playback options.
- Google Map (3):** A map showing the location of Kyneton, Kilmore, and Wallan.



extension within industrial FUI project




LIMAH: (description and) browsing



Hyper Media Analytics v0.0.1-SNAPSHOT

+ Add to favorites
★ My favorites
Account ▾

May 24, 2015



ABOUT THIS ARTICLE

Type: PRESSE
 Publisher: Le Monde
 Category: Article
 Link: Le Monde.fr
 Published on: May 24, 2015
 Author:
 Related documents: 14

WORD TAG

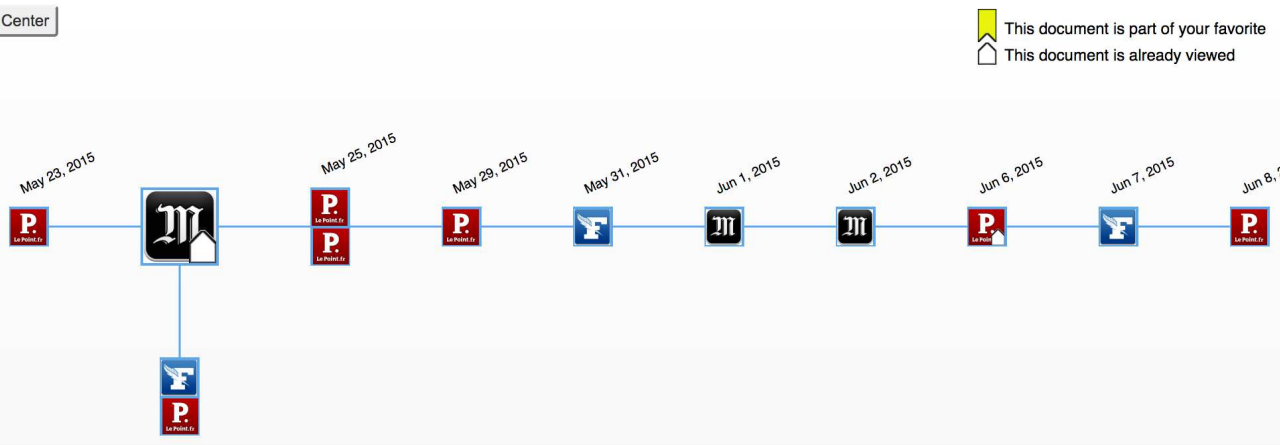
17
4 ministre
3 Grèce
3 Voutsis
2 chemin
2 créanciers
2 Tsipras
2 Alexis
2 spectre
2 Varoufakis

NAMED ENTITIES

Nikos Voutsis
 Grèce
 Grèce
 monétaire international
 FMI
 Tsipras
 BBC
 Yanis Varoufakis
 Grèce
 M. Varoufakis

Center

This document is part of your favorite
This document is already viewed



La Grèce souffle le chaud et le froid sur ses remboursements

La posture officielle du gouvernement de gauche radicale est d'annoncer qu'on servira d'abord les pensions et les retraites, puis les créanciers.

Nikos Voutsis, le ministre de l'intérieur de la Grèce, a déclaré à la chaîne de télévision Mega, dimanche 24 mai, que la Grèce n'avait pas d'argent pour payer le Fonds monétaire international (FMI) en juin.

Aucun porte-parole d'Alexis Tsipras, le premier ministre, n'a commenté ces propos. Sur la BBC, Yanis Varoufakis, le ministre des finances, a uniquement souligné que la Grèce avait fait « *un pas énorme* » dans la négociation d'un accord avec ses créanciers internationaux pour éviter la faillite. Il a qualifié de « *catastrophique* » pour son pays l'idée de quitter la zone euro. « *C'est maintenant aux institutions de faire leur part. Nous les avons rejointes aux trois quarts du chemin, elles doivent nous rejoindre sur un quart du chemin* », a déclaré le ministre.

M. Varoufakis a aussi dit au *New York Times* cette semaine :

Le spectre de caisses entièrement vides

FILTER GRAPH

PRESSE

- Le Figaro
- Huffington Post
- Le Point
- Liberation
- Le Monde

VIDEO

- France Television


RADIO

- RMC
- Radio France

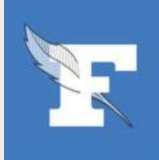
TWITTER

- Twitter


RELATED ARTICLES



Michel Sapin : une sortie de la zone euro serait "une catastrophe pour la Grèce"



La Grèce ne remboursera pas le FMI en juin



La Grèce incapable de rembourser le FMI en juin

Probabilistic models are (almost) everywhere

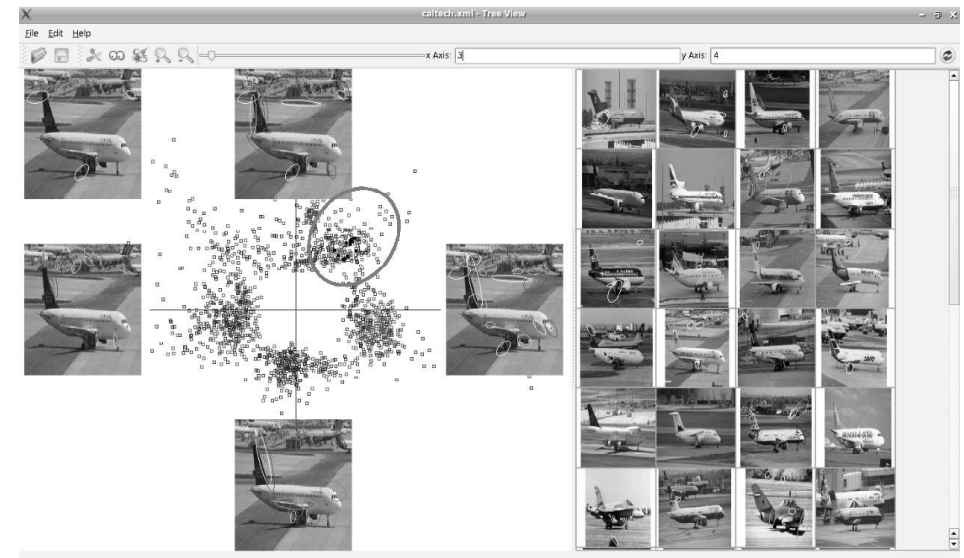
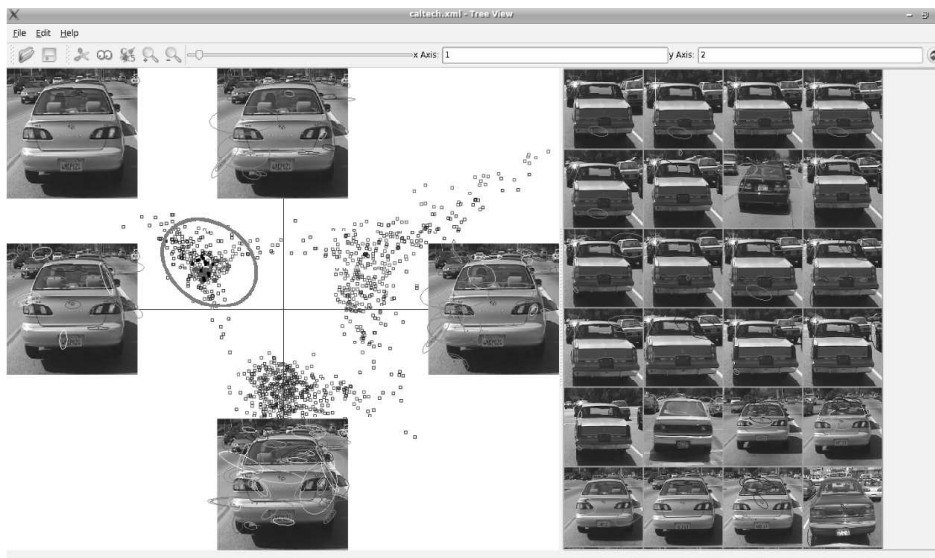
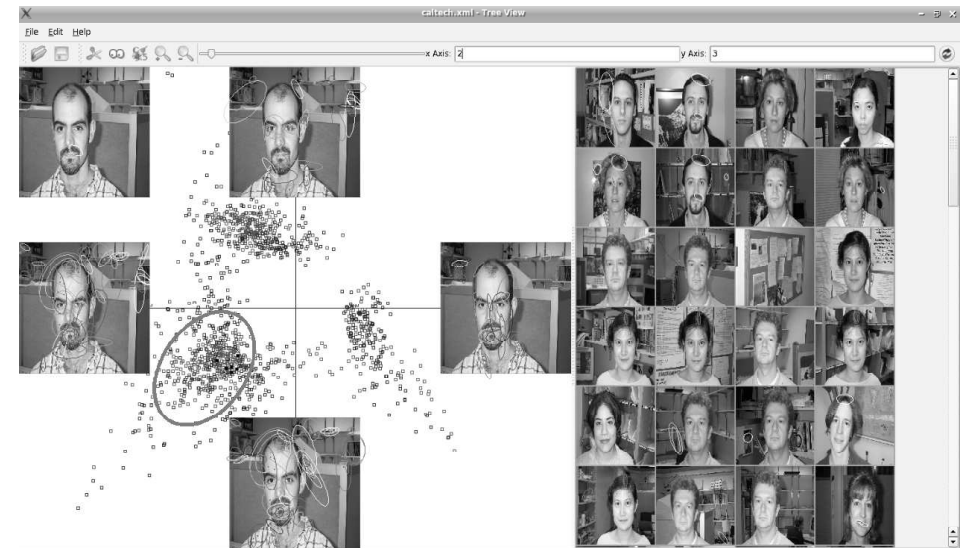
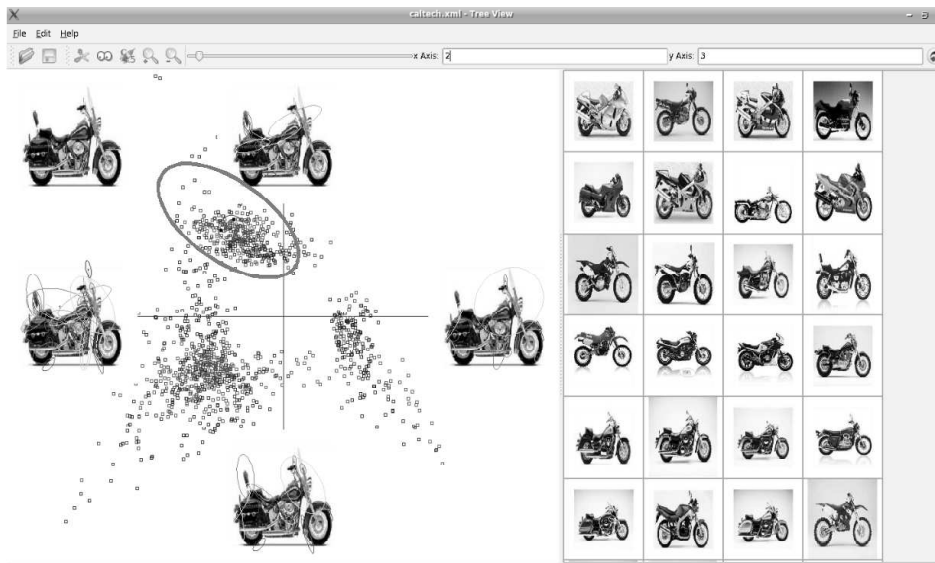
- Image indexing and description
 - *Fisher vector, image/tag co-occurrence analysis, PCA, CCA, FCA, etc.*
- Speech recognition
 - *hidden Markov acoustic models, n-gram language models*
- Keyword extraction, named entity recognition
 - *term frequency analysis, conditional random fields*
- Topic segmentation
 - *mutual information (IM^3), language model, burst detection, etc.*
- Content matching
 - *term frequency, latent Dirichlet allocation, cross correlation analysis, etc.*
- Collection organization, navigation
 - *data visualization, graph analysis, etc.*



Deep learning now replacing good old probabilistic models

→ Is deep learning statistics? If not, can both be combined?

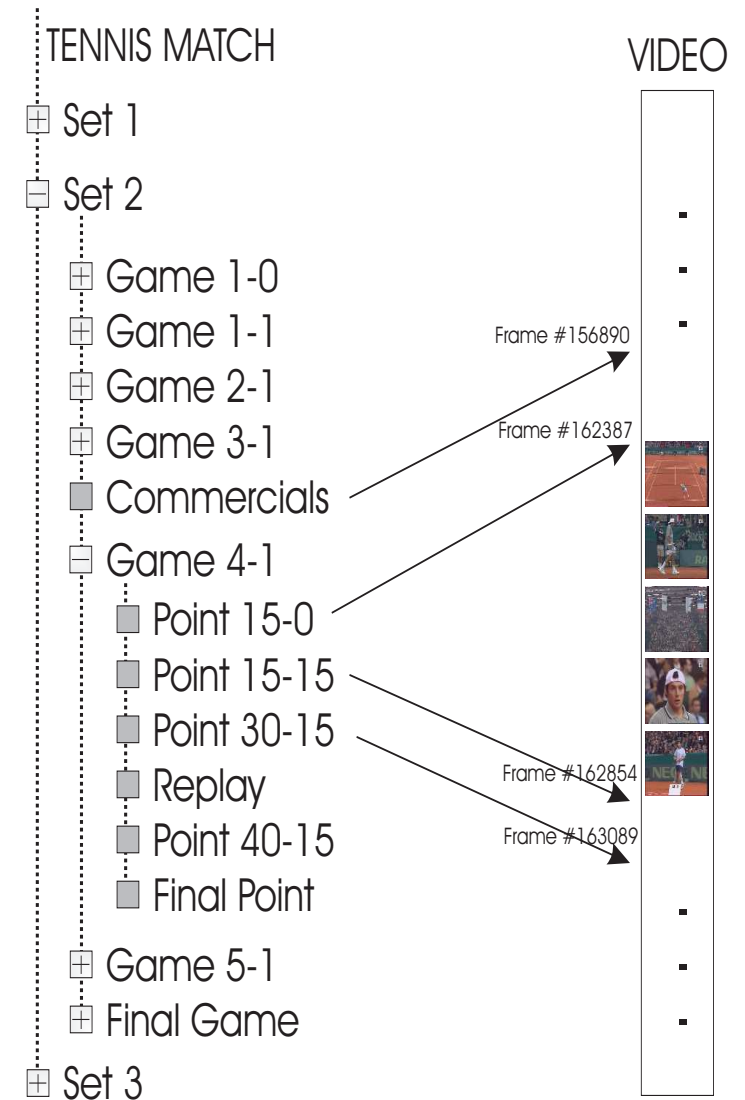
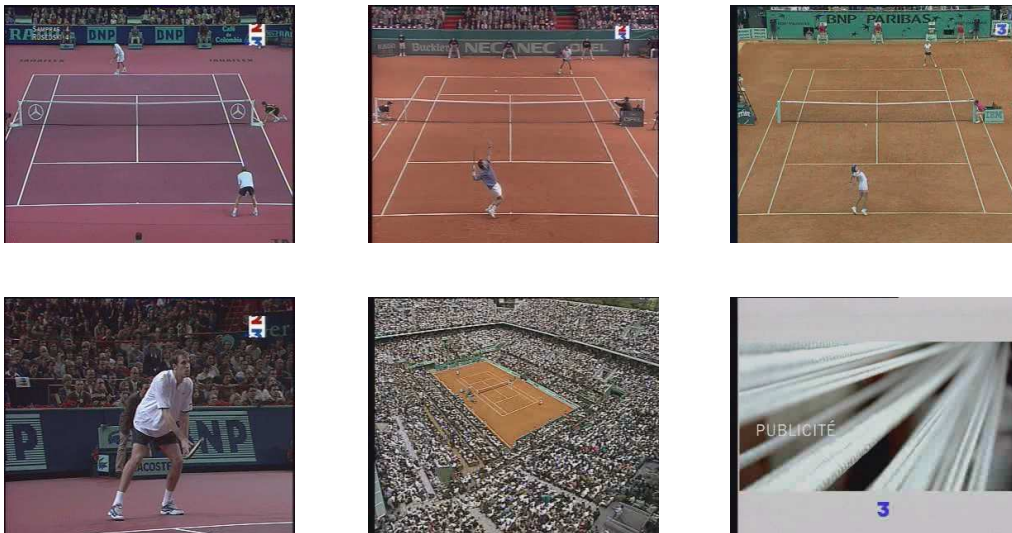
Example: images and factorial correspondence analysis



[Nguyen-Khang Pham *et al.*. Intensive use of factorial correspondence analysis for large scale content-based image retrieval. *Advances in Knowledge Discovery and Management*, 2010]

Another example: HMM for video structuring

- four generic scenes
 - missed serve + rallye, rallye, replay, other
- *prior* knowledge
 - editing rules, tennis rules
- *posterior* knowledge
 - images & soundtrack

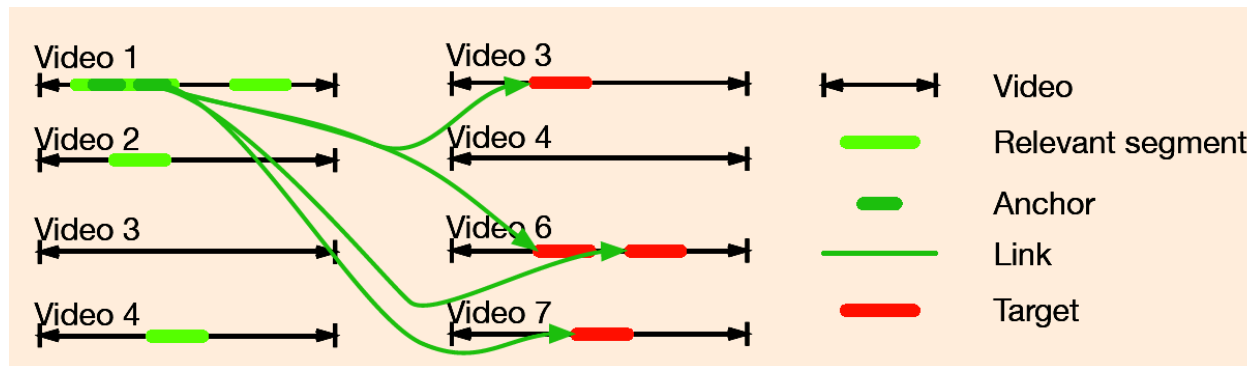


[Ewa Kijak *et al.*. Audiovisual integration for tennis broadcast structuring. *Multimedia Tools and Applications*, 2006]

What in media hyperlinking?

The Mediaeval/TRECVID hyperlinking task scenario:

1. anchor detection: finding potential anchors in the videos
2. fragment linking: creating a ranked list of segments related to the anchors



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

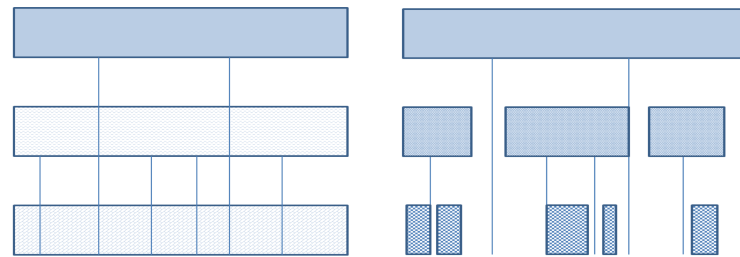
anchoring + hyperlinking = organizing a collection for analytics



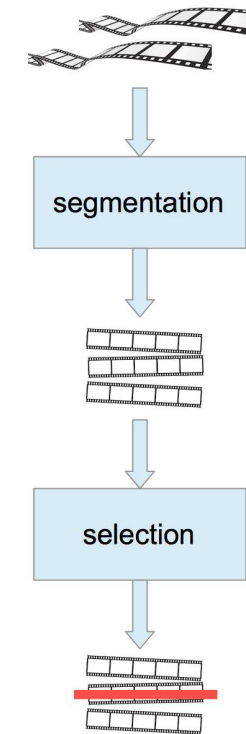
A two step language-based hyperlinking framework

Topic segmentation based on transcripts

- ASR-adapted linear topic segmentation
[Guinaudeau *et al.*, Mediaeval 12; Şimon *et al.*, EMNLP 13]
- hierarchical topic segmentation
[Guinaudeau *et al.*, Mediaeval 13; Şimon *et al.*, SLAM 13]
- fragmentation



[Şimon *et al.*, RANLP 15]

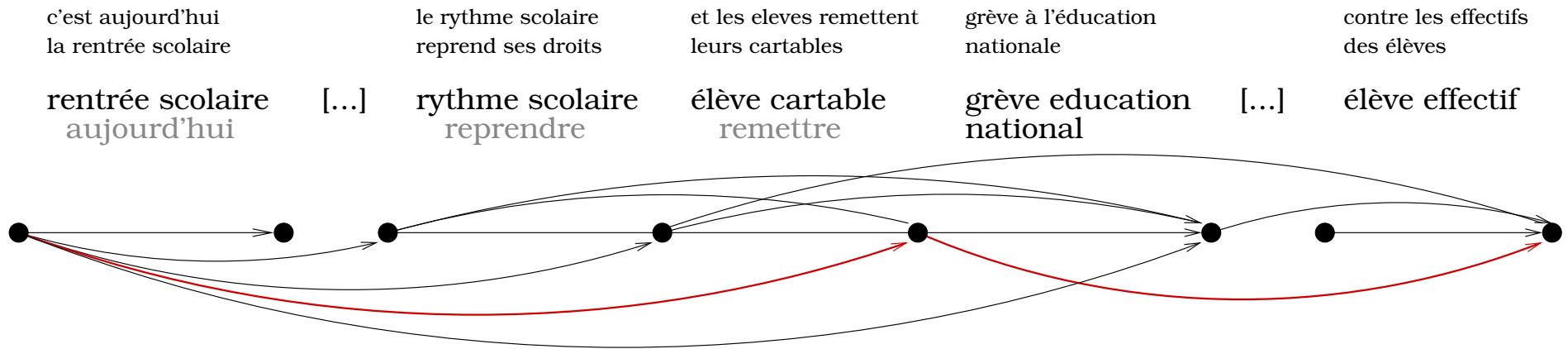


Standard language-based ranking

system	reference			ASR		
	lin+bow	lin+ngram	hie+bow	lin+bow	lin+ngram	hie+bow
P	0.31	0.42	0.26	0.20	0.33	0.19
P_{tol}	0.25	0.41	0.26	0.14	0.30	0.17

Mediaeval 2013 [Şimon *et al.*, SLAM 14]

Graph-based topic segmentation



vertex values = lexical cohesion = generalized probability

$$\hat{s} = \operatorname{argmax}_{s_1^m} \sum_{i=1}^m \ln(P[w_{a_i}^{b_i} | S_i]) - \alpha \ln(n)$$

$$\Delta_i = \left\{ P_i(u) = \frac{C_i(u) + 1}{z_i}, \forall u \in V_K \right\} \quad \ln P[S_i; \Delta_i] = \sum_{j=1}^{n_i} \ln P[w_j^i; \Delta_i]$$

Graph-based topic segmentation (cont'd)

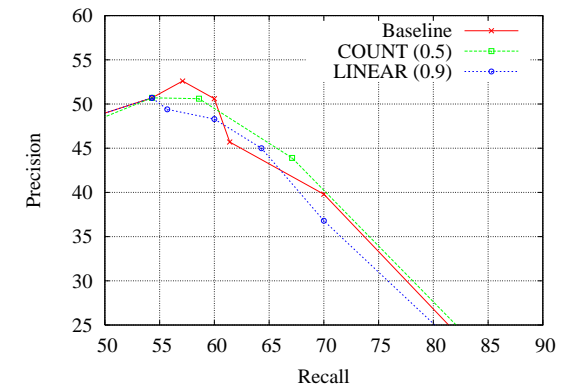
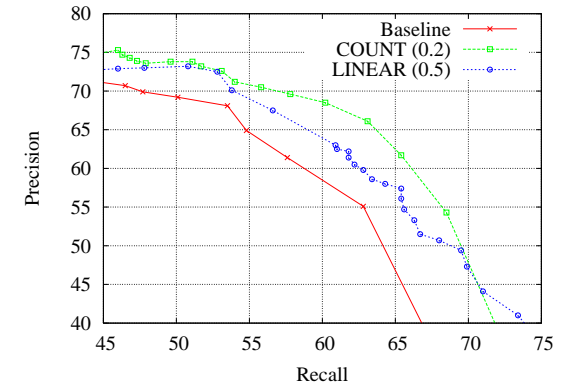
Improving language modeling

- using confidence measures
- adding semantic relations to cohesion

$$C_i''(u) = C_i(u) + \sum_{j=1, w_j^i \neq u}^{n_i} r(w_j^i, u)$$

- using interpolated language models

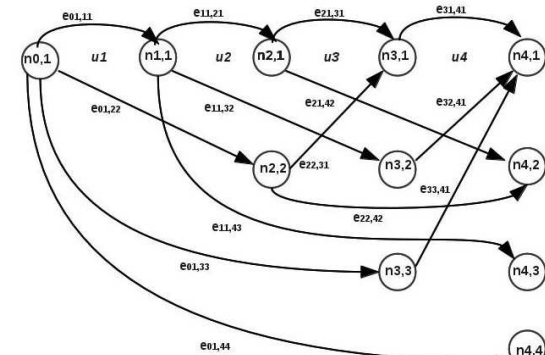
$$\ln P[S_i; S_i, T] = \sum_{j=1}^{n_i} \ln(\lambda P[w_j^i; \Delta_i] + (1-\lambda)P[w_j^i; \Delta_t])$$



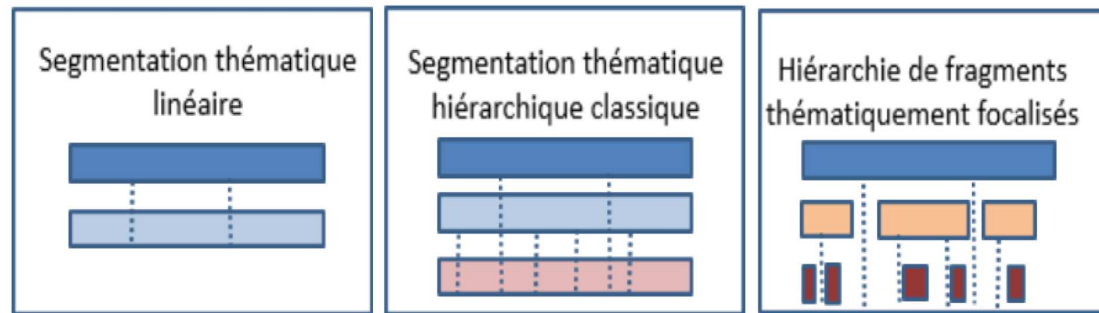
Remove independence assumption

$$P[W|S_1^m] = P[W|S_1] \prod_{i=2}^m P[W|S_i, S_{i-1}]$$

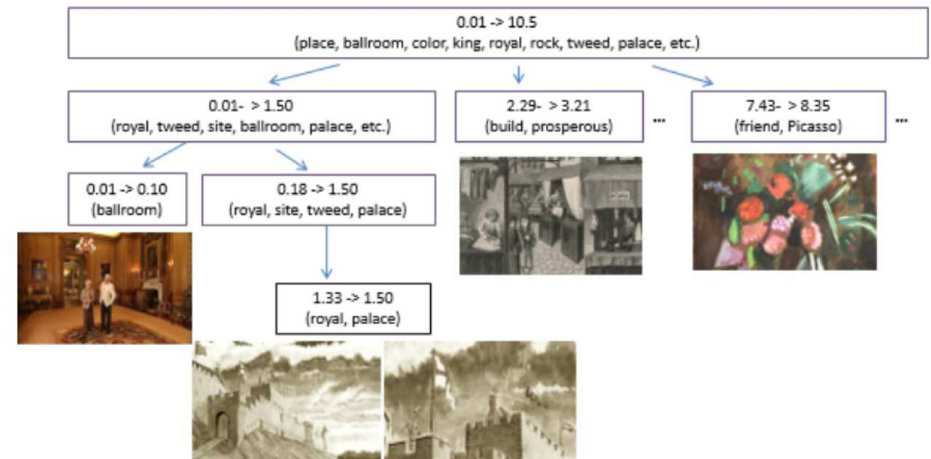
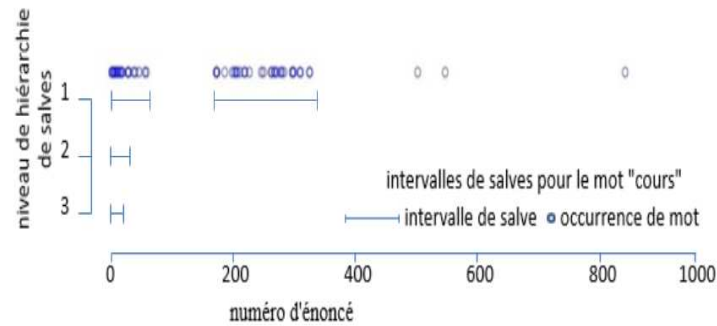
where $\ln P[W|S_i, S_{i-1}] = \ln P[W_i|S_i] - \lambda \Delta(W_i, W_{i-1})$



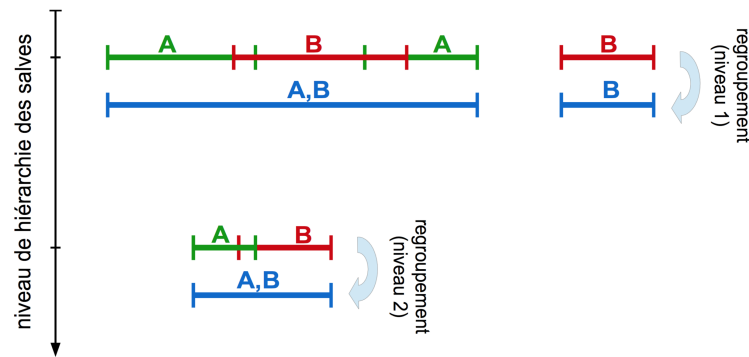
From segmentation to fragmentation



From burst analysis ...



... to topic fragments



Topic models for content matching

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]

Hierarchy with 10 levels, trained independently on BBC collection transcripts

- level 1, $K_1 = 50$, broad topics z_i^1 ($i \in [1, K_1]$)
- level 10, $K_{10} = 1,700$, fine-grain topics z_i^{10}

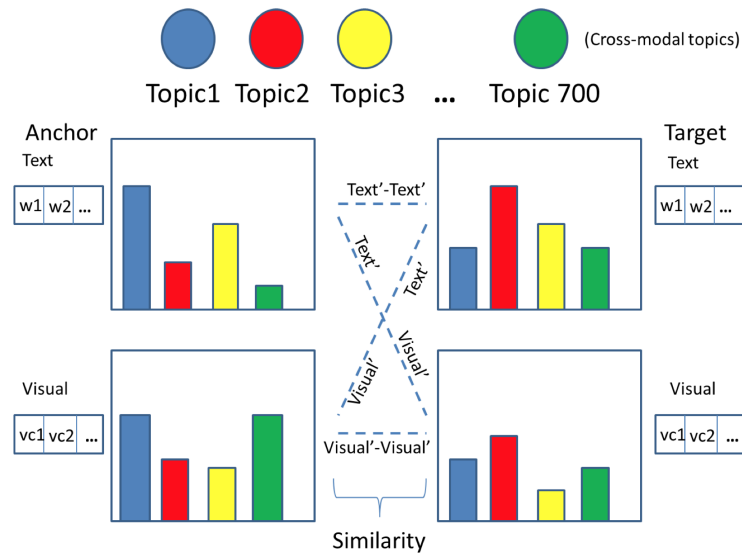
year	number of topics (K)				
	direct	50	150	300	700
2013	0.25	0.44	0.34	0.35	0.34
2014	0.19	0.18	0.25	0.26	0.21

Target reranking task, using targets from all MediaEval participants: relevance after 15 s of the top-10 targets [Simon *et al.*, SLAM 15]

[combination into hierarchical stuff goes here]

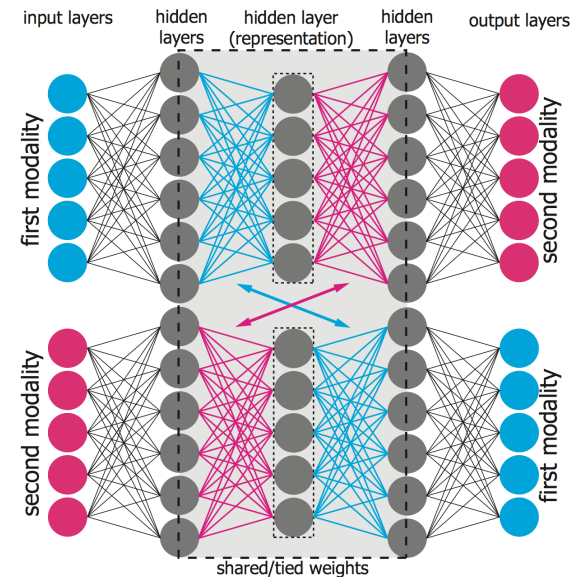
Multimodality to improve diversity

Bimodal extension of LDA mapping words and visual concepts via topic-specific distributions



[Bois *et al.*, TRECVID 15]

Crossmodal matching with symmetrical bi-directional auto-encoders

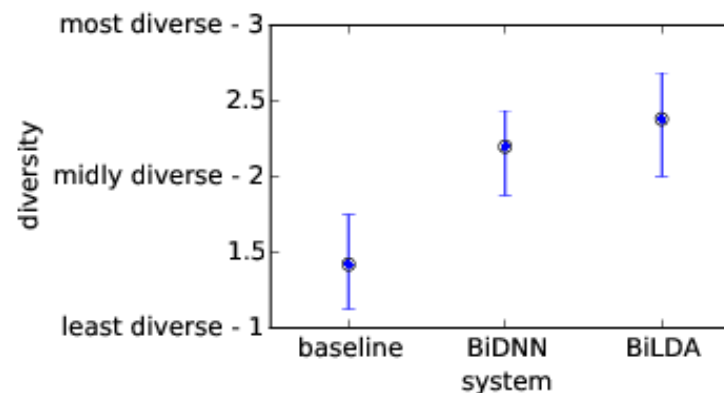


[Vukotić *et al.*, ICMR 16]

Multimodality to improve diversity (cont'd)

Measuring diversity of the top-5 relevant targets found by each approach

	transcripts			concepts		
	n_u	\bar{d}_a	\bar{d}_i	n_u	\bar{d}_a	\bar{d}_i
baseline	29.8	0.51	0.61	35.6	0.61	0.71
BiDNN	40.8	0.20	0.12	46.7	0.42	0.31
BiLDA	40.0	0.25	0.16	38.0	0.48	0.41



Intrinsic measures of diversity

Perceived diversity (25 subjects)

[Bois *et al.*, submitted to MMM 17]

Legend:

n_u = number of unique key words/concepts in top-5 targets

\bar{d}_i = average similarity among the top-5 targets

\bar{d}_a = average similarity between anchor and top-5 targets

Acknowledgements

Many thanks to all who contributed to the results presented here (in alphabetical order): Laurent Amsaleg, Rémi Bois, Morgan Bréhinier, Sébastien Champion, Vincent Claveau, Camille Guinaudeau, Ewa Kijak, Sien Moens, Pascale Sébillot, Ronan Sicre, Anca-Roxana Şimon, Arnaud Touboulic, and Vedran Vukotić.

And very likely others that I might have forgotten.

Work presented here benefited from the financial support of: BPI France (Quaero), CominLabs & ANR (LIMAH), EIT ICT Labs (OpenSEM) and Région Bretagne (ARED).