MediaEval Benchmark

University Toulouse III - Paul Sabatier

MediaEval Challenge

Team Karadoc

[M2 CMI SID]

25th November, 2016

Professors :

Isabelle Ferrané
Jérôme Farinas
Julien Pinquier
Lynda Tamine-Lechani
José Moreno

Students :

Salima Azzou
Axel Bellec
Max Halford
Joseph Meunier
Giovanni Zanitti

General contents

Table of contents

The MediaEval challenge?

- Evaluation of Multimedia Access and Recovery **Algorithms**
- Predicting the main **theme** of a video
- 3 weeks work, 20 hours per week

Team coordination

- Each member mostly worked on a **single media**
- We all **wrote** the report and we **organised** tasks as a group

Organisation and tools

- **Python** - Programming language
- **GitHub** - Git repository hosting service
- **Travis** - Continuous integration tool
- **Slack** - Real-time messaging and task-planning for teams
- **ShareLaTeX** - Real-time collaborative redaction tool

Table of contents

Text

- Metadata
  - `duration`
  - `licence`
  - `size`
  - `title`
  - `uploader_id`
  - `uploader_login`
- Instance Matrices
  - Terms in the description
  - Terms in the title
  - Assigned tags

Audio

- Transcript
  - Speakers and their attributes per file (`time, gender`)
    - «Features» extraction (`entropy, number of male, number of female, frequency`)
  - Transcript instance matrix (with confidence percentage)

Audio (2)

- Audio Signal : video to audio conversion made with `ffmpeg`
  - Mel Frequency Cepstral Coefficients (MFCCs)
  - Energy vectors

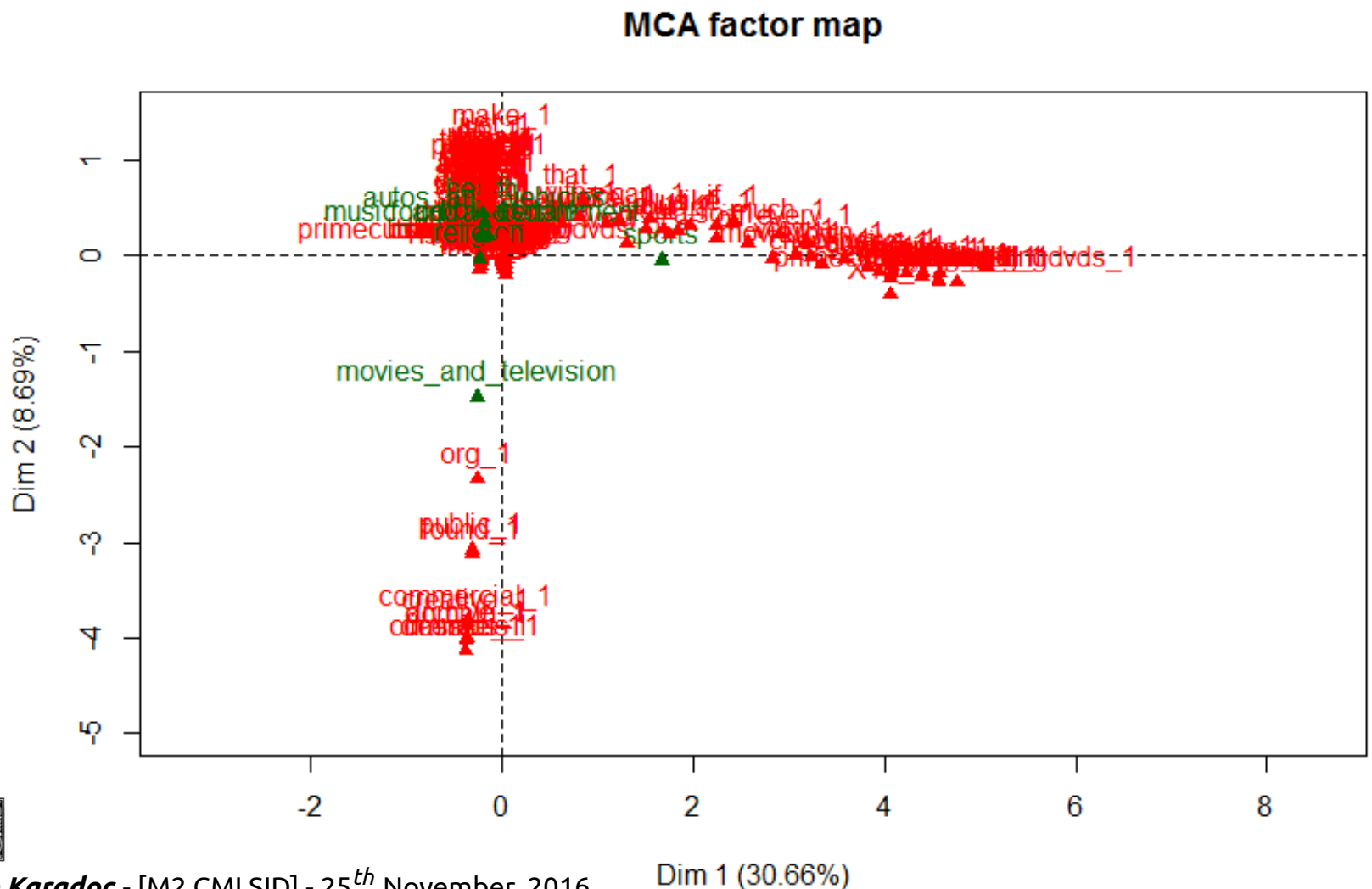
Video

Figure: Sample keyframe (or shot)

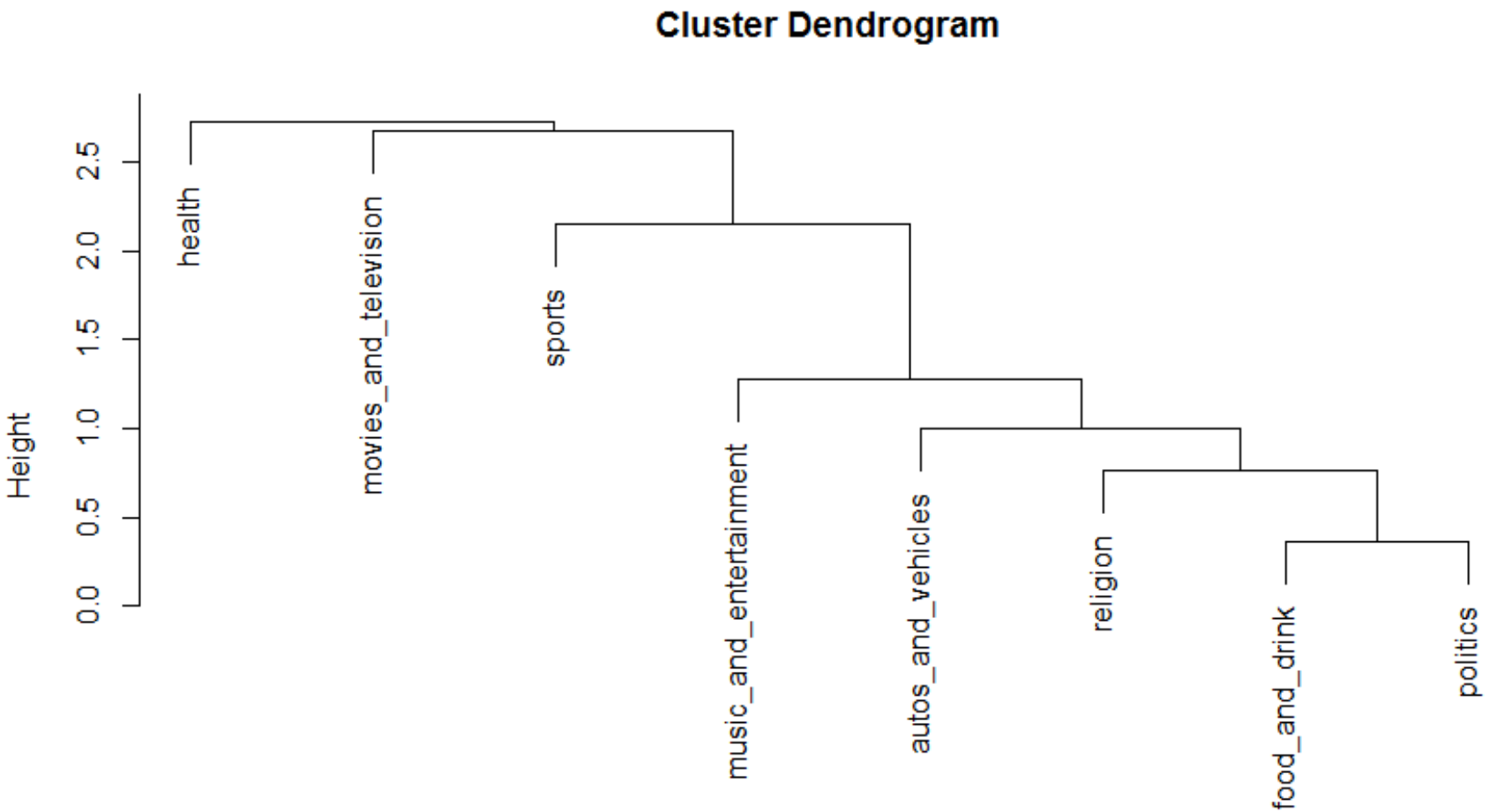
- Shots
  - Number of keyframes per video
  - Color histograms
  - Image segmentation block by block followed by averaging
  - Facial recognition
  - Optical Character Recognition (OCR)

Table of contents

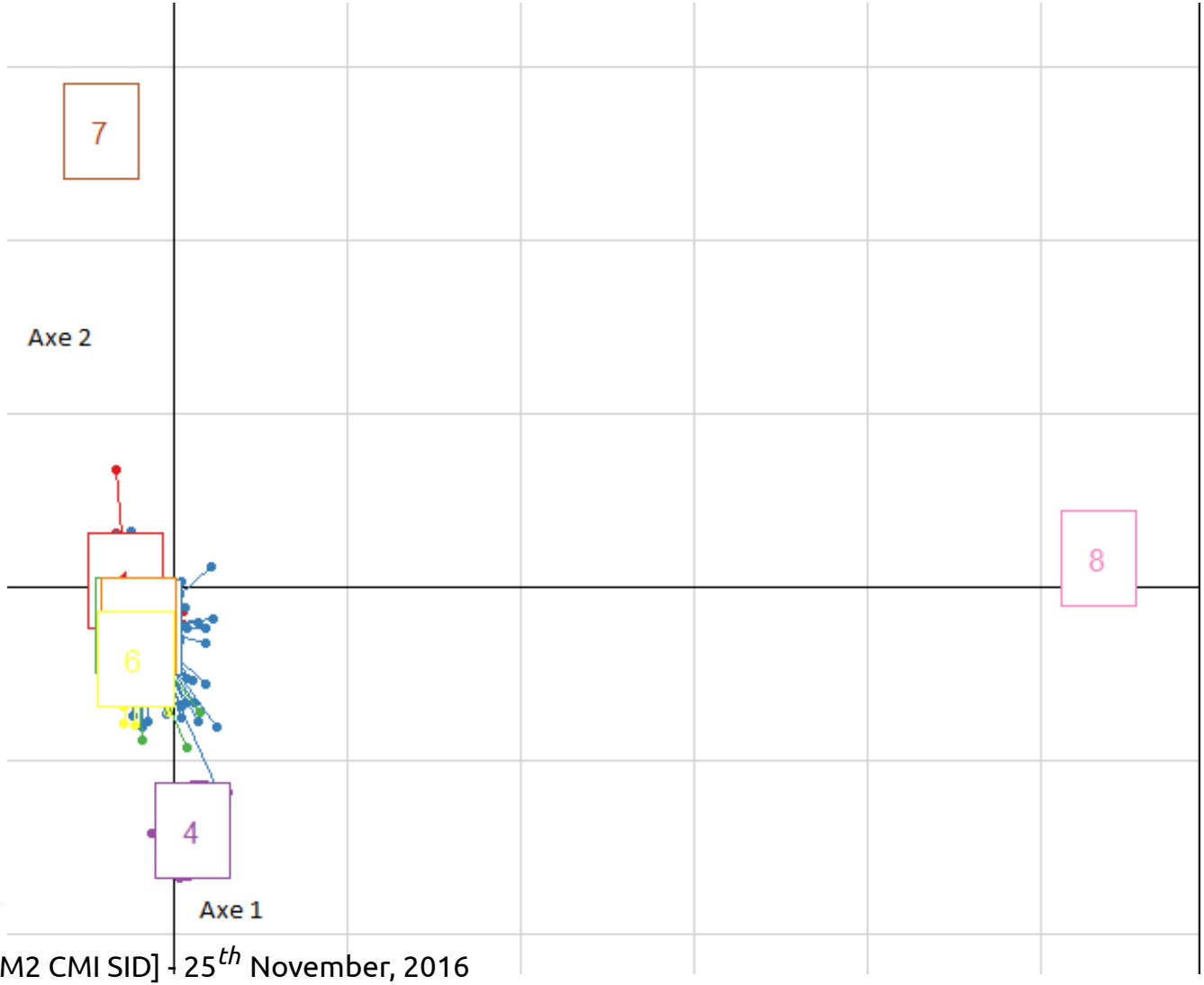
ACM


MCA factor map
make_1
that_1
movies_and_television
org_1
public_1
found_1
sports
dvds_1
Dim 2 (8.69%)
Dim 1 (30.66%)
1
0
-1
-2
-3
-4
-5
-2
0
2
4
6
8

ACM


Cluster Dendrogram
Height
0.0
0.5
1.0
1.5
2.0
2.5
health
movies_and_television
sports
music_and_entertainment
autos_and_vehicles
religion
food_and_drink
politics

ACM



Summary

	key	descri	titre	initiale
sports	23	25	29	63
food_and_drink	404	244	415	46
music_and_entertainment	41	7	33	135
health	12	186	15	73
politics	15	24	6	141
autos_and_vehicles	8	8	34	8
religion	47	42	29	45
movies_and_television	25	39	14	64

Figure: Number of documents assignated to each category

Other clustering methods

- K-means clustering
- Multidimensional scaling (MDS)
- Hierarchical document clustering
- Latent Dirichlet Allocation
- Non-negative Matrix Factorization

Table of contents

Main idea

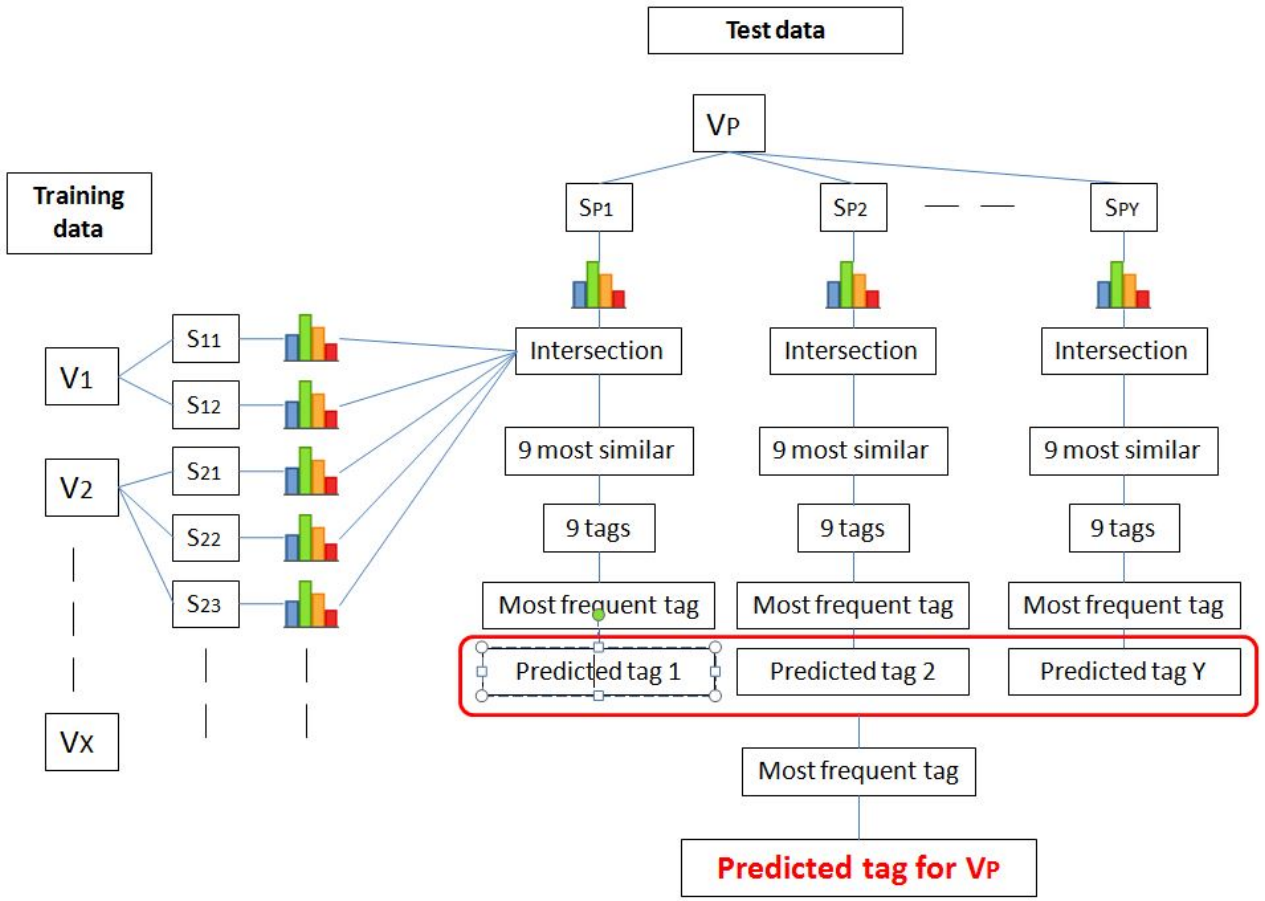
- We wanted to use independent classifiers right from the start
- We had experience working on different themes from our lab exercices
- We thought we could “merge” them in the end with a “metaclassifier”

Intuitions

- Tags and description seem to be quite general
- More features than documents: linearly separable space
- Some videos don't have a lot of action: shot analysis
- Some videos only have a single speaker

KNN on color histograms

- Use the algorithm built in class

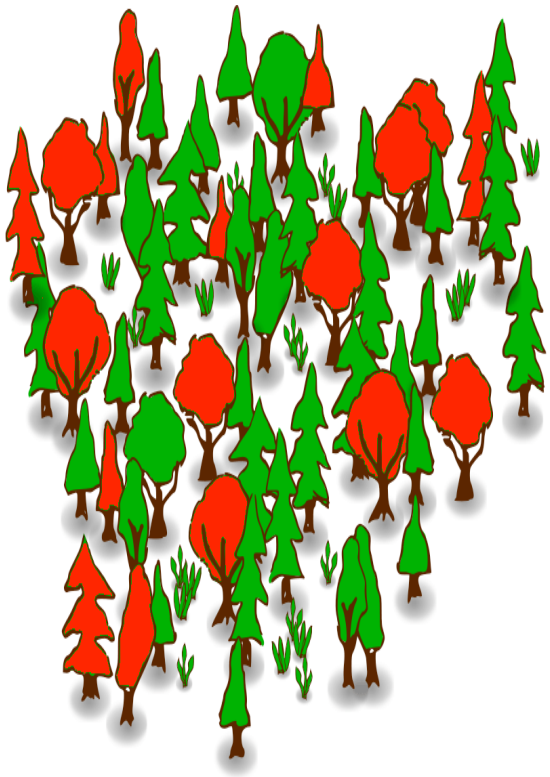


Text classification

- SVM works very well on a linearly separable space
- Naïve Bayes not as good even through bagging
- Lemming and stemming help a lot to generalize
- We could have went further by using a thesaurus

Random Forest classification

Applied to global features



- Shots
  - has_text
  - nb_faces_max
  - nb_shot
- Speakers
  - entropy
  - nb_M
  - nb_F
  - Freq_M/F
- Metadata
  - duration
  - size
  - uploader_id

Table of contents

Scoring

	Precision	Mean training time
Naïve Bayes on metadata TF	0.839	0.697
Top terms on metadata TF	0.682	0.356
KNN on metadata TF	0.822	0.285
Linear SVM on metadata TF-IDF	0.879	0.287
Naïve Bayes on transcription TF	0.548	0.372
Top terms on transcription TF	0.354	0.313
KNN on trans TF	0.654	0.294
Linear SVM on trans TF-IDF	0.665	0.269
Random Forest on speakers and shots features	0.661	0.241
Neural network on signal energy	0.249	0.231

Table: Classifiers score

Metaclassification

- Store the predictions for each classifier
- Matrix with as many rows as documents and as many columns as classifiers
- Run a Random Forest on this matrix hoping to gain accuracy

Table of contents

Difficulties

Technical difficulties

- Library dependencies under Windows OS:
  - `scikit-learn` (machine-learning library)
  - `opencv3` (`python` wrapper of `opencv` library)
- Hard to stay focused and creative for 3 weeks :)

Personal gain

- We learned how to code with `python`
- We perfected our **team working** skills
- We used new tools
- We used and developped our **statistical** skills

Thanks for listening