EURECOM
Sophia Antipolis

Zero-Shot Classification of Events for Character-Centric Video Summarization

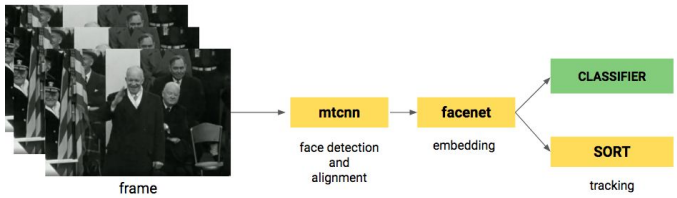
Alison Reboud,
Ismail Harrando,
Pasquale Lisena,
Raphaël Troncy

EURECOM Team - Takeaways

- **Method**: Event classification driven and character-centered approach
- **Contribution**: Showing that using a list of named important events is an efficient starting point for this task, even matching results obtained on the 2020 edition with enriched with external data
- **Results:** 37.6% for the best run across characters (34.6% for subtask)
- **Observation**: Contrary to our 2020 approaches runs with more shots did not always yield significantly better results / Performance is close for both the main task and subtask

Last Year's Approach

- **Scraping synopses from Fandom EastEnders Wiki[1]**
  Hypothesis: every sentence from these summaries represents an important event to be added to the final video summary
- **Shots Selection with Face Recognition[2] for query characters :**
  - Crawl the web with name of actor+ "EastEnders"
  - Perform face detection and recognition
- **Matching Scenes and Content**
  - BoW representation for each scene / sentence from the summary
  - Pick the scenes with the highest matching score



[1] https://eastenders.fandom.com/wiki/EastEndersWiki

[2] https://github.com/D2KLab/Face-Celebrity-Recognition

EURECOM

Approach: Events Zero Shot Classification

Life events labels and their perceived likelihood (scale from 1 to 5)

Label	Weight
extramarital affair	1.98
get divorced	1.96
illegitimate child	1.45
institutionalized for emotional problem	1.43
happily married	4.05
serious accident	2.96
murdered	1.81
attempt suicide	1.26
blackmailed	1.86
unfaithful spouse	2.23
sexually assaulted	2.60
abortion	1.41

Code:
https://github.com/MeMAD-project/trecvid-vsum

Eastenders videos + subtitles

Face recognition for query characters

Align subtitles with video shots

ZSC with events labels

Scenes selection based on ZSC scores and event likelihood

* only keeping scenes adjacent to query characters

Life Events / Soap Opera

Table 3.3: Life events labels and their perceived likelihood (scale from 1 to 5)

Label	Weight
extramarital affair	1.98
get divorced	1.96
illegitimate child	1.45
institutionalized for emotional problem	1.43
happily married	4.05
serious accident	2.96
murdered	1.81
attempt suicide	1.26
blackmailed	1.86
unfaithful spouse	2.23
sexually assaulted	2.60
abortion	1.41

[1] Gayle Seese. Soap opera viewers' perceptions of the real world. Master's thesis, University of Central Florida, 1987

Matching and runs generation

- **Generating an 'importance' score**

$$score(shot\_i) = max\_l \in labels(zsc(trans\_i, l) * weight(l) * log(len(trans\_i))$$

where *shot i* is the unique id of the shot, *trans_i* is its corresponding transcript, *labels* is the list of events, with their importance expressed with *weight*

- **Order shot by importance:** only keep best match (across labels) for each shot
- **Choose the N most matching shots in chronological order,** cutting the longest shots if the time restriction is not respected

Results

Overall results

Team	Main task	Subtask
ADAPT	**30.15%**	17.25%
EURECOM	29.55%	**30.10%**
NII_UIT	18%	29.85%

- Results close to those obtained last year with an approach relying on fan written synopsis
- Queries not easy to answer as they require to understand the relation with other characters than those of interest

Detailed results for the queries about Archie with 20 shots included in the summary

Query	Main task	Subtask
What happens when Phil throws Archie in to a pit?	Yes	No
What happens after Danielle reveals to Archie that Ronnie is her mother?	Yes	No
Where do Peggy and Archie get married?	No	No
What happens when Archie arrives at the pub after Peggy invited him?	No	No
What happens when Archie is kidnapped?	Yes	No

Average score for each run

TeamRun	Score
Eurecom1	17.4%
Eurecom2	30.4%
Eurecom3	32.8%
Eurecom4	37.6%

Subtask - Answering Questions

- Same problem, different approach
- Question-Answering: the goal is to predict where the answer to the question lies in the text (content subtitles).
- For that, we use HuggingFace's Transformer QA pipeline (Using longformer as a base model, pretrained on Squad-v2 QA task) ⇒ Keeping the highest scoring answers for each question
  - The model is not trained on this kind of task (the answer is usually included in the options as an affirmation, not a line of dialog)

Subtask results

TeamRun	Score
Eurecom1	32.2%
Eurecom2	31.8%
Eurecom3	30.8%
Eurecom4	34.6%

Lessons Learned

- Zero Shot classification is a promising direction for generic character-centric summarization
- Aligning shots and subtitles => not straightforward
- Any narrative-level understanding of character roles can be better suited for shot selection instead of character presence in the scene (e.g. character interactions/involvement in subplots of other characters)
- Sound: maybe in the future?

Thank you for your attention!

Any questions?

MeMAD

Methods for Managing Audiovisual Data

INSTITUT CARNOT Télécom & Société numérique

EURECOM