TRECVID 2018
Social media video story linking

João Magalhães[1], David Semedo[1], Saverio Blasi[2]

[1]Universidade NOVA de Lisboa, [2]BBC

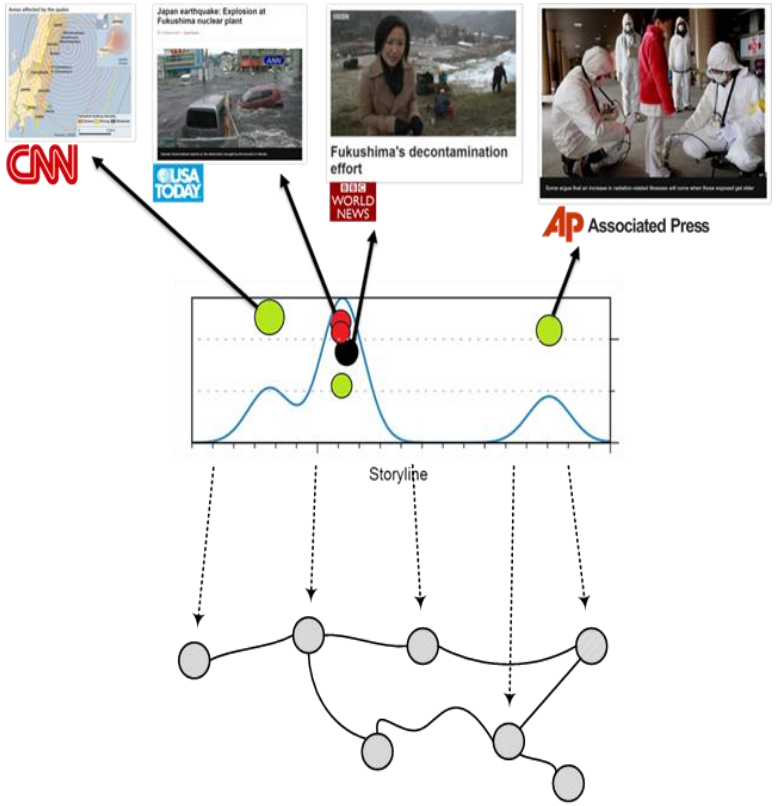
NIST, Gaithersburgh

14 November 2018

Context

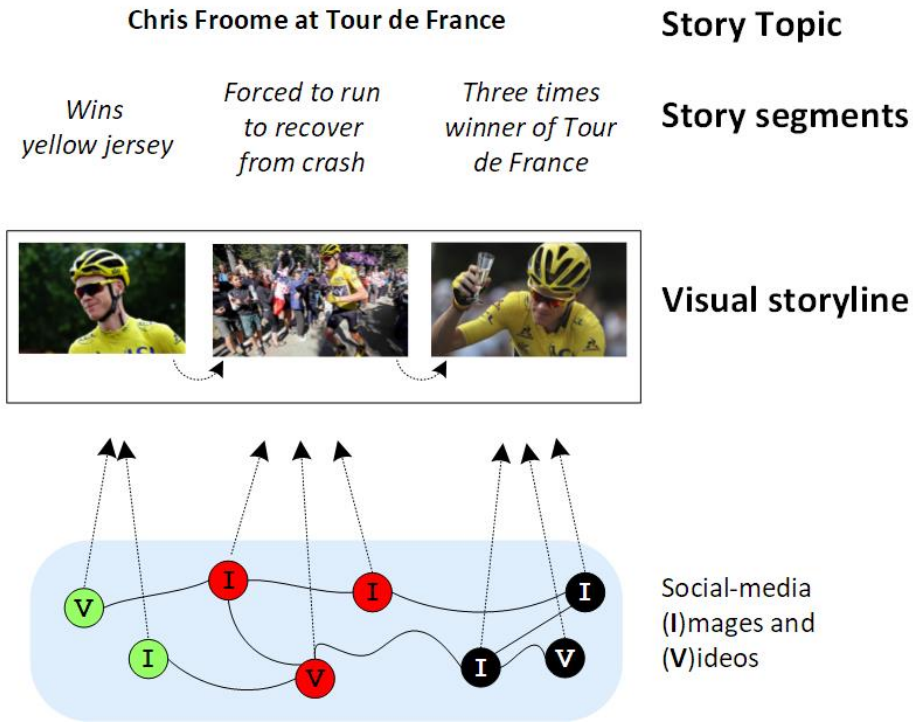
- The goal is to illustrate a news story with social media content.

- Starting from a news story topic and a stream of social media video and images, the goal is to link a story-segment to image and video material, while also preserving a good flow of the whole visual story.



Task

- Topics correspond to a sequence of story segments
- A story segment is a sentence (+ an image) query with some a strong visual component
- Systems should retrieve the video and image that satisfy the two requirements:
  - Best illustrates the news segment;
  - Makes the best transition from the previous video/image illustration.

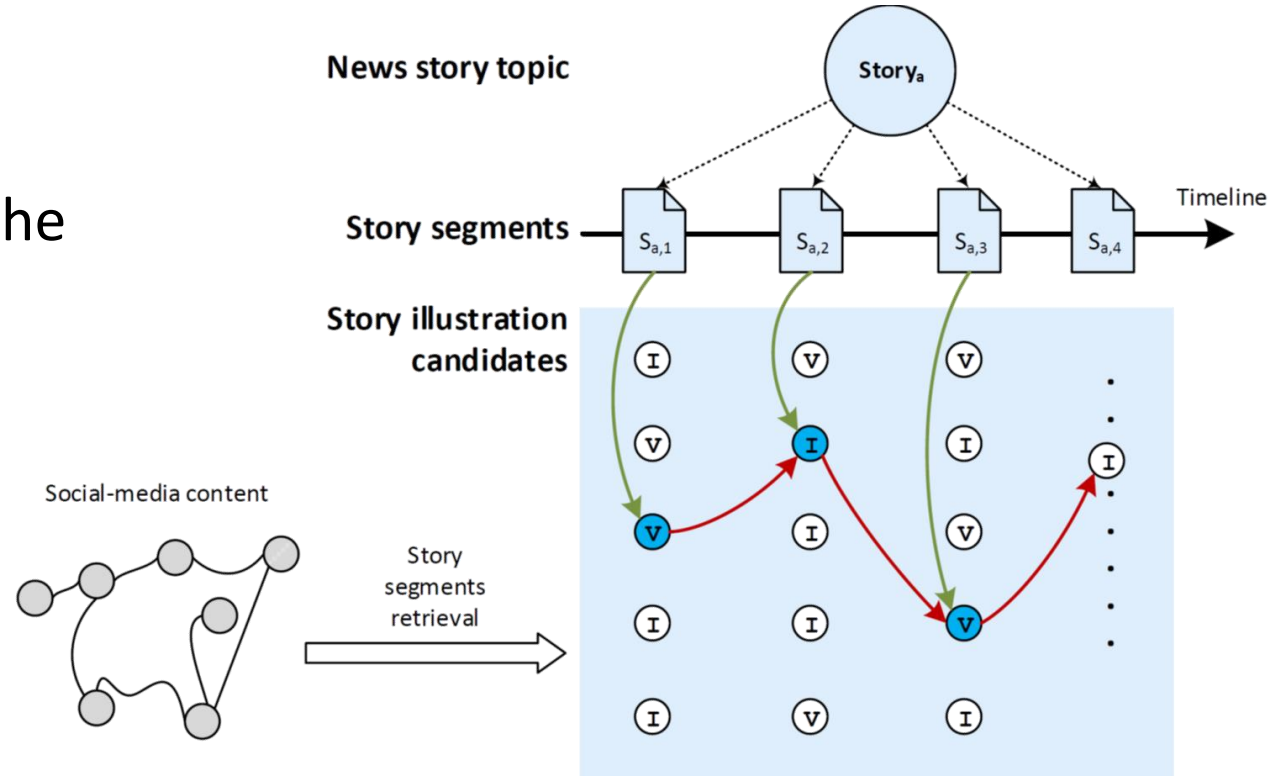


Who is the end user?

- Broadcasters
- Social media
- News press
- Instituions/companies involved in the coverage of live events.

Task and evaluation

- A possible angle of attach is to first retrieve a large number of videos per segment and then determine the right sequence.
- This rationale was used to evaluate the output.
- Other approaches can consider the entire dataset.



Quality metric

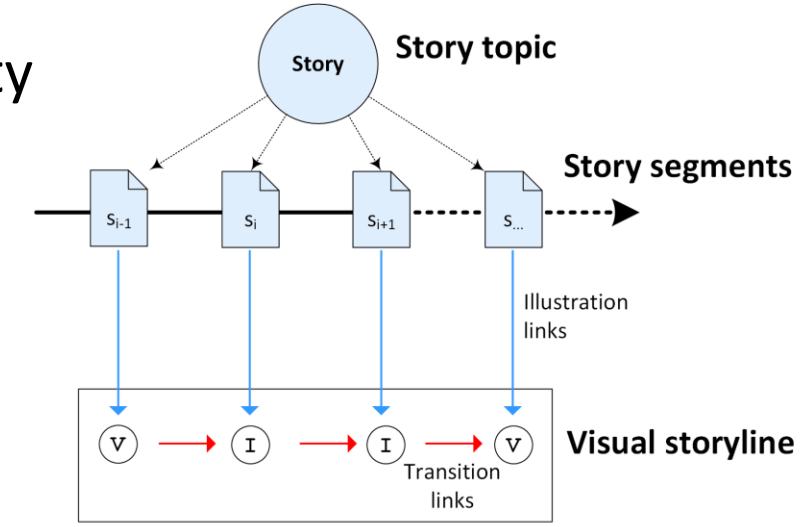
- The quality metric is used to evaluate the overall quality of a story in

$$Quality = \frac{1}{N}\sum_{i=1}^{n} pairwiseQuality(i, i-1)$$

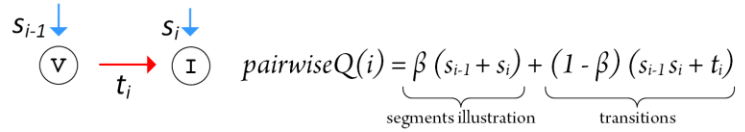
$$pairwiseQuality(i, i-1) = 0.4 \cdot (s_{i-1} + s_i) + 0.2 \cdot (t_i + s_{i-1} \cdot s_i,)$$

  - The first parcel addresses the relevance of the illustrations.
  - The second parcel addresses the transitions.

- This creates relevance judgments for the relevance of media and validity of the sequence.



(a) Visual story editing assessment framework.



(b) Visual story quality assessment metric.

Events

- To run the Social Media Visual Storytelling Linking task, we considered news stories about two events:
    - **Edinburgh Festival:** The event has a duration of 3 weeks in August.
    - **Le Tour de France:** The event has a duration of 23 days in July.

Dataset

- Data sources:
  - **Twitter images and videos:**
    - Edinburgh Festival: over 32k images and 6.2k videos;
    - Le Tour de France: over 66k images and 19k videos.
  - **Flickr images:**
    - Edinburgh Festival: over 10k images;
    - Le Tour de France: over 11k images.
- Data split:
  - **Training**: 2016 data
    - Groundtruth only for images
    - >Story topics and relevance judgments
  - **Test**: 2017 data (only story topics)

Topics

- Topics:
  - Driven by existing data alone
  - Driven by news articles in online media
- Aimed to have always 4 query segments per story.
- All query segments had only text.
- Some stories had multimodal query segments from BBC, but due to delays in copyright licensing, we had to drop these topics.


```
"story_title": "What is the EdFest?",
"story_id": 101,
"segments": [
  {
  "segment_id": 1,
  "text": "Music shows",
  "keywords": "Music shows"
  },
  {
  "segment_id": 2,
  "text": "Theater and comedy",
  "keywords": "Theater and comedy"
  },
  {
  "segment_id": 3,
  "text": "Circus",
  "keywords": "Circus"
  },
  {
  "segment_id": 4,
  "text": "Street performances",
  "keywords": "Street performances"
  }
]
```

Teams

- 13 teams registered for the LNK task.
  - (6 Eur, 6 Asia, 1 US)
- 7 teams downloaded data.
- Out of these, we received two submissions:
  - (one team only solved the search by relevance part and did not submit)
  - ADAPT, IE
  - NOVA Search, PT

Relevance judgments

- Stories were judged in a real storyteller prototype.

- Users judged:
  - Relevance of the segment (judged as 0, 1 and 2)
  - Quality of the transition (judged as 0, 1 and 2)
  - Rated the story as a whole (judged between 0 and 5)

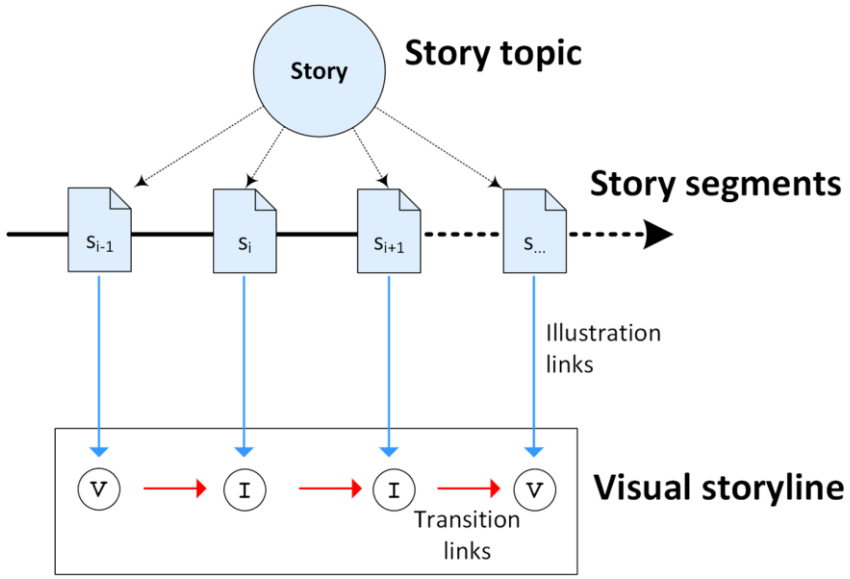
- Some runs had segments of 30 mins!!!

Annotators guidelines

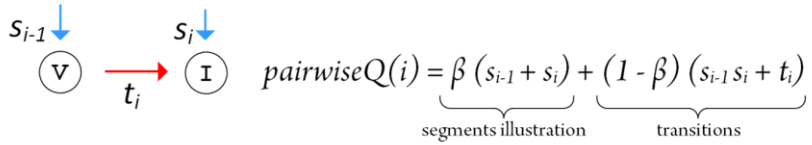
- **Relevance of the visual illustration to story segment description (blue links)**:
  - 0 (no relation between the segment illustrations),
  - 1 (visual or semantic relation between the two segments),
  - 2 (strong visual or semantic relation between the two segments).

- **Transition consistency of the illustration to pairs of story segments (red links)**:
  - 0 (no relation between the segment illustrations),
  - 1 (visual or semantic relation between the two segments),
  - 2 (strong visual or semantic relation between the two segments).

(Videos were judged positively if any part of it was relevant.)

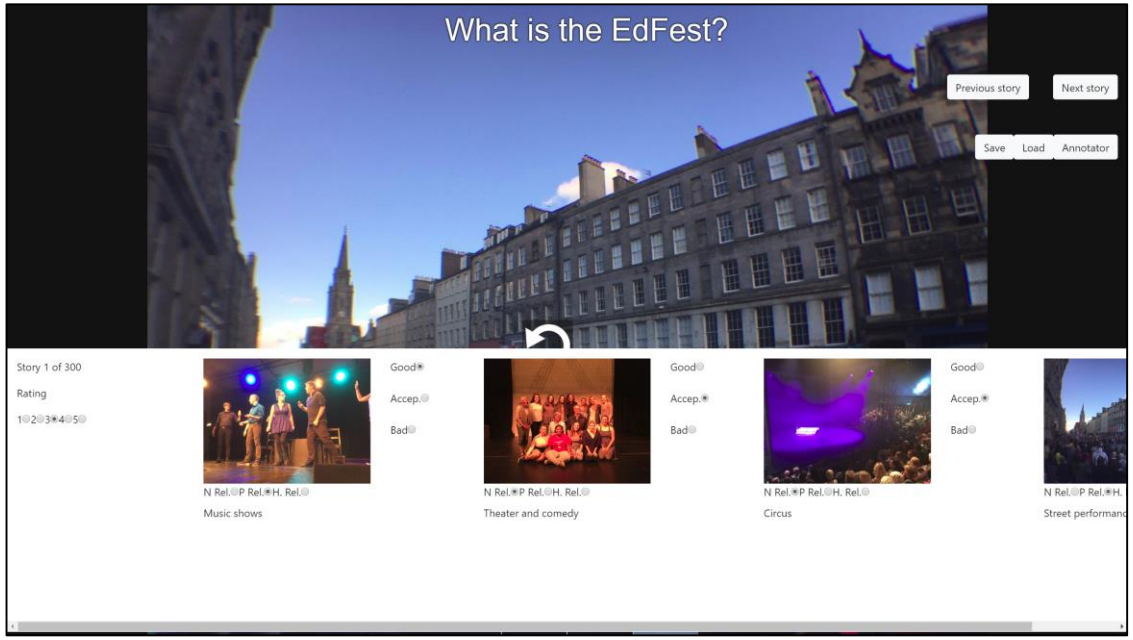


(a) Visual story editing assessment framework.



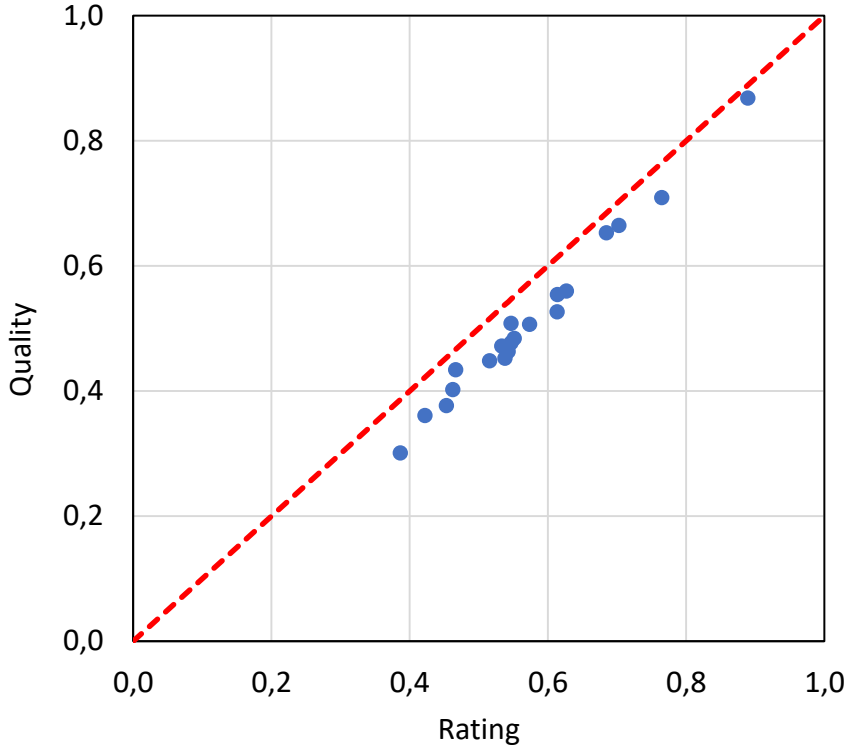
(b) Visual story quality assessment metric.

Storyteller annotator prototype



Quality vs rating correlation

- Metric achieves a 0.98 correlation with human judgments (ratings)
- The raking of the runs change only slightly between the computed quality and the judged quality
- The parameters of the quality metric can be adjusted when we get enough data.



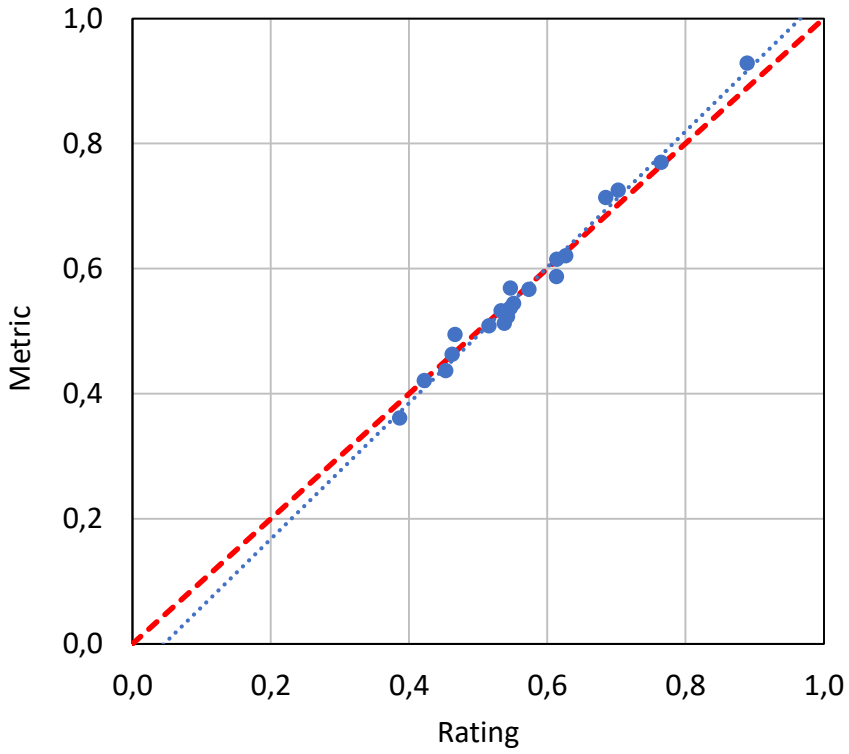
After correcting the average of the metric

- Correcting the intercept is not enough
  - The slope also needs to be corrected

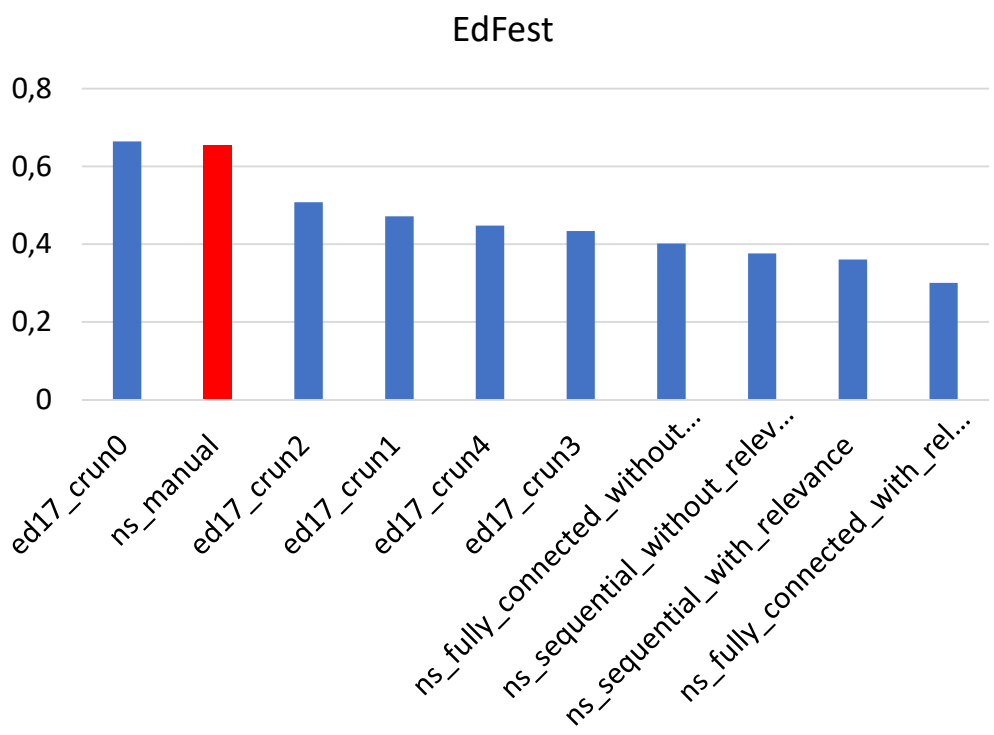
- However, the used metric offered a very good correlation the overall judged quality

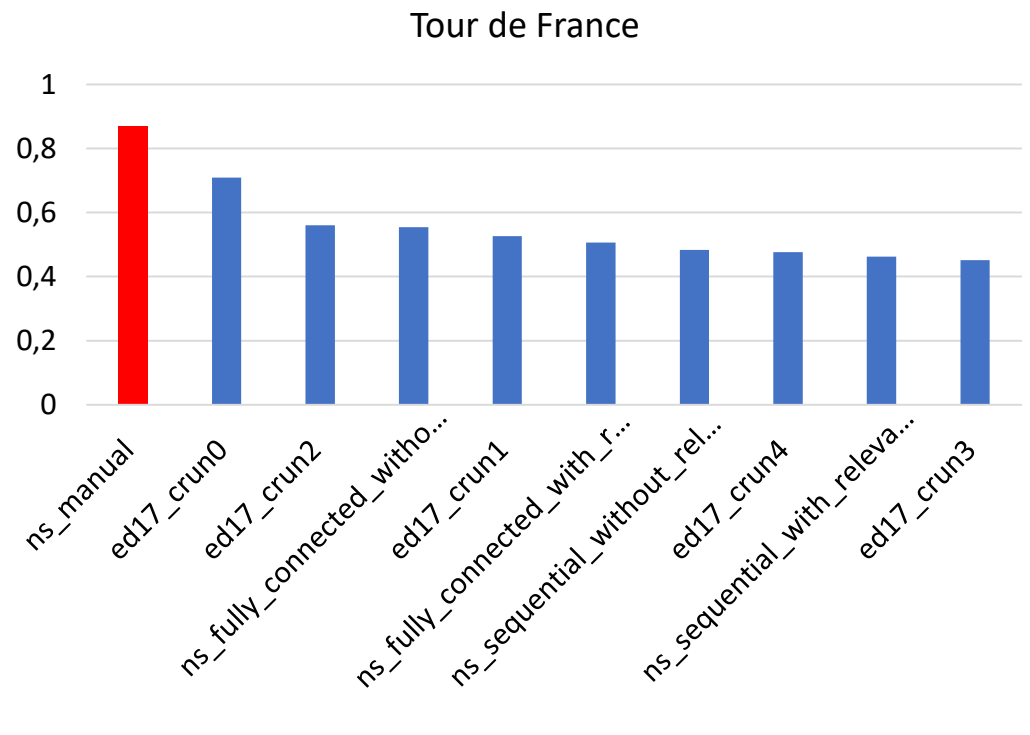
$$Quality = \frac{1}{N}\sum_{i=1}^{n} pairwiseQuality(i, i-1)$$

$$pairwiseQuality(i, i-1) = 0.4 \cdot (s_{i-1} + s_i) + 0.2 \cdot (t_i + s_{i-1} \cdot s_i,)$$

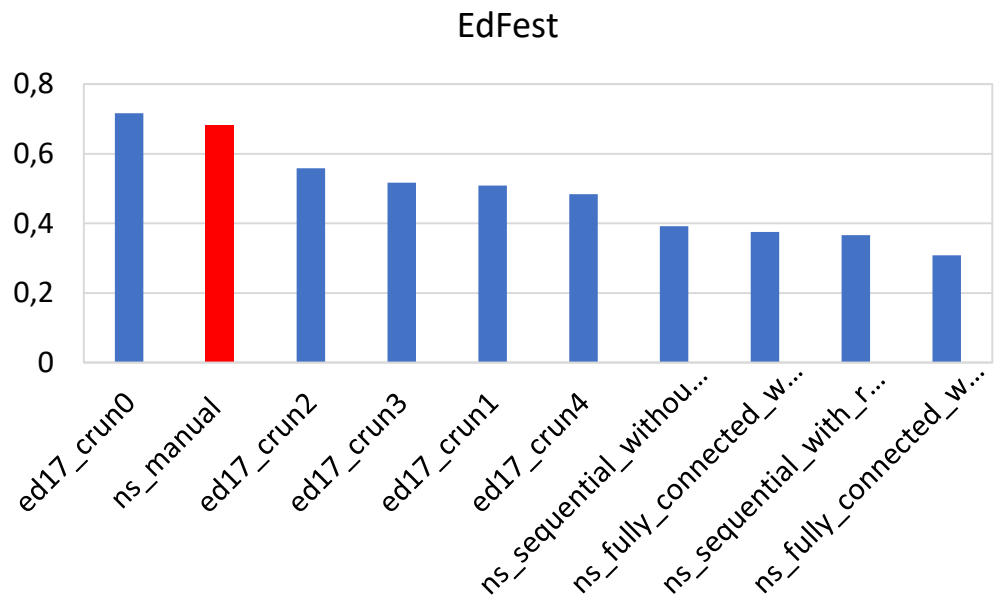


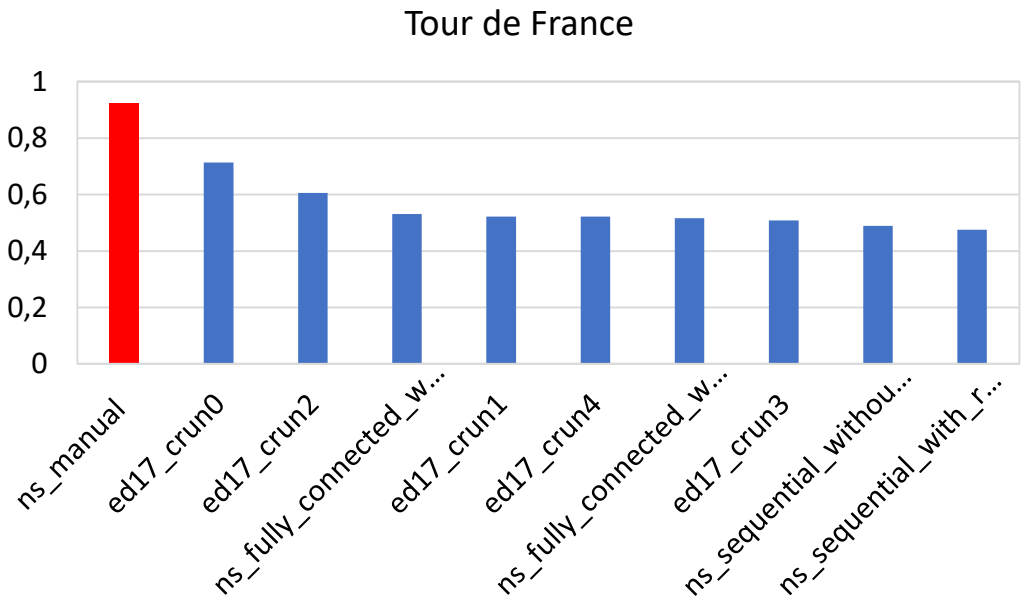
Runs results



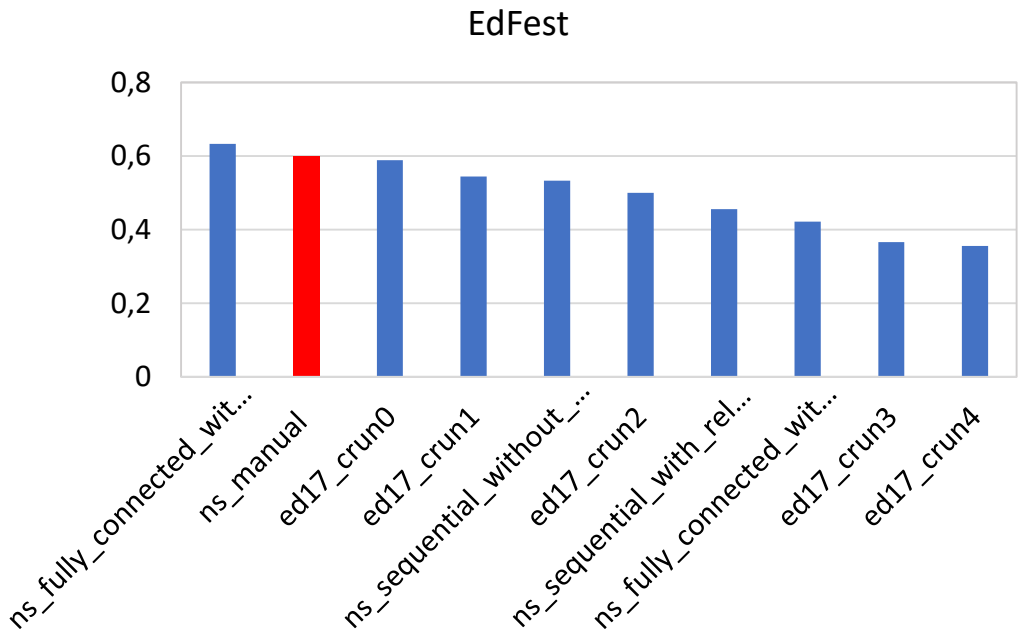


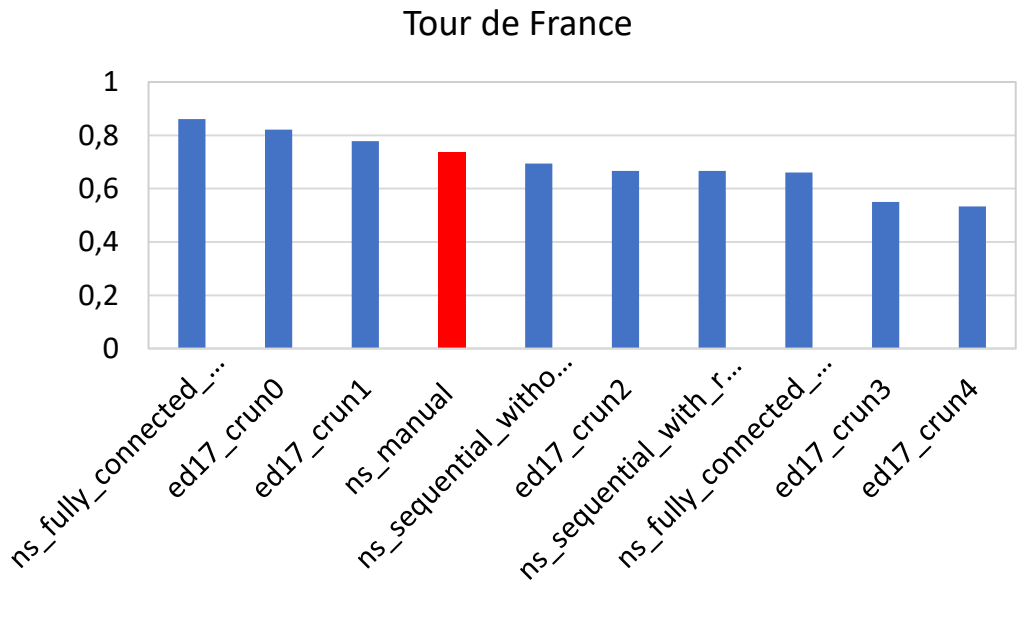
Relevance of story segments





Transitions





Takeaway messages

- Relevance is critical for the story quality.
  - Video semantic search is key.
- Transitions are also important, but to a lesser extent.
- Videos need to be croped to a well defined interval.
- Relevance judgments are strongly correlated with users' general perception of story quality.

Acknowledgements

Plans for 2019

- Dataset:
  - Use the same dataset with new stories?
  - Access to Twitter content will be always a bottleneck. However, it makes it a lot more interesting!
  - Include the Flickr content.

- Task:
  - Stories are fixed in segments. Should it be variable?
  - Limit video segments to 30 secs.
  - Correlate the story length with the story quality.
  - Allow stories with multiple candidate segments? (too complex to evaluate)

- Attract more teams!