Analyzing non-linear video usage in an introductory x-MOOC about basic linear algebra

J. Triay[1], J. Minguillón[1], T. Sancho-Vinuesa[1], V. Daza[2]

[1]Universitat Oberta de Catalunya, Barcelona, Spain
[2]Universitat Pompeu Fabra, Barcelona, Spain

MOOCs

- First MOOC in 2008
- 2011 Sebastian Thrun, Peter Norvig, Artificial Intelligence, 160.000 paticipants.
- Very popular in 2012
- Huge MOOC platforms
  - edX
  - MiriadaX
- From individual experiences to institutional
  - Future Learn
  - UCATx (Based on Open edX)

“Decoding Algebra” course

- Introductory course for university students.
- 5 weeks, one module / week, 3-5 hours / week
- Two kind of resources:
  - 97 videos ranging from 5' to 18’.
    - Main resource.
    - Theory, exercises, challenges.
  - Quizzes with 8 or 10 questions.
    - Three attempts.
    - Best attempt.

Motivation

- Exploratory analysis.
  - Video consumption.
  - Hot spot and bottlenecks detection.
  - Correct storyboard.
  - Number of videos.

- Build a set of tools for analyzing UCATx courses.

Data gathering and preprocessing

- 194 course participants

- Around 450.000 events in six weeks
  - EdX tracking logs format. (.JSON)
  - Video events: *play_video*, *seek_video*.
  - *Seek_video*, jump during the reproduction.
  - Python scripts.
  - One file with all video jumps: ID, start time, final time.

Data analysis (I)

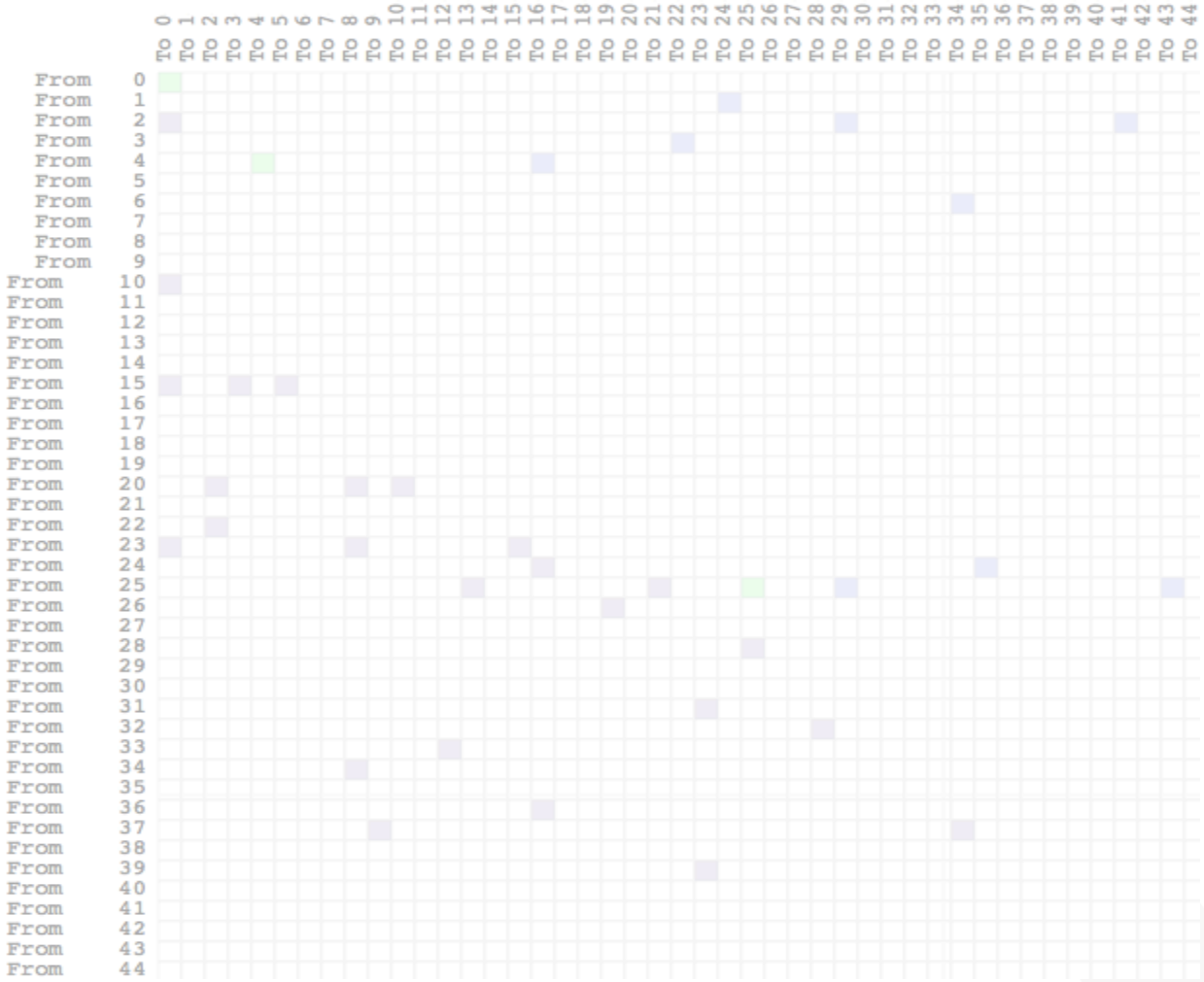
- Heatmaps D3.js
  - Visualize the previous .csv file with all jumps.
  - Blue color gradient, forward jumps.
  - Red color gradient, backward jumps.
  - Darker color, more jumps; white color, no jumps.
  - Y-axis, where participants begin the jump.
  - X-axis, where participants finish the jump.

Data analysis (II)

- One second resolution
  - Square matrix
  - Each cell represents a second where the participant begins the movement of the play bar and the second where finishes the movement.
  - Number of cells, equal to the square of the video length in seconds.
- Different problems for long videos:
  - Lot of time to render.
  - Most of heatmap is white.

Data analysis (III)

- One second resolution image.



Data analysis (VI)

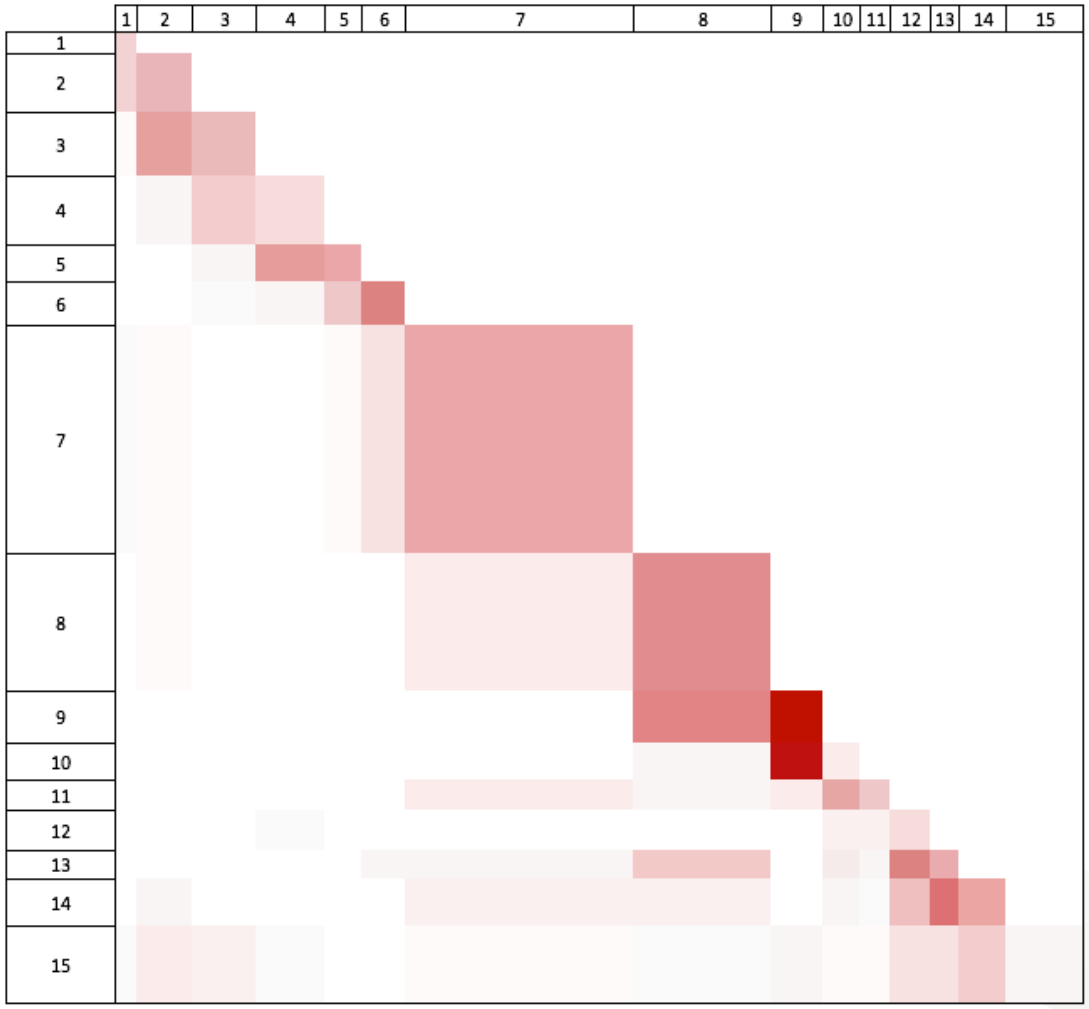
- We build different heatmaps with different cell resolution 2-20 seconds.
  - Different problems.
    - 1-10 seconds resolution:
      - Long videos, same as the 1 resolution.
    - 11-20 seconds resolution:
      - Short videos, cause distortion.
      - Jumps are very close, in the same cells.

Data analysis (V)

- We developed heatmaps depending on the internal structure (storyboard).
- Analyse the internal structure of the video
  - Different scenes.
- Now each cell represent a scene.
- It remains being a square matrix.
- Cell size is scaled according scene duration.
- Normalization applied.
- Solved previous problems of data sparseness, renderization time and close jumps.

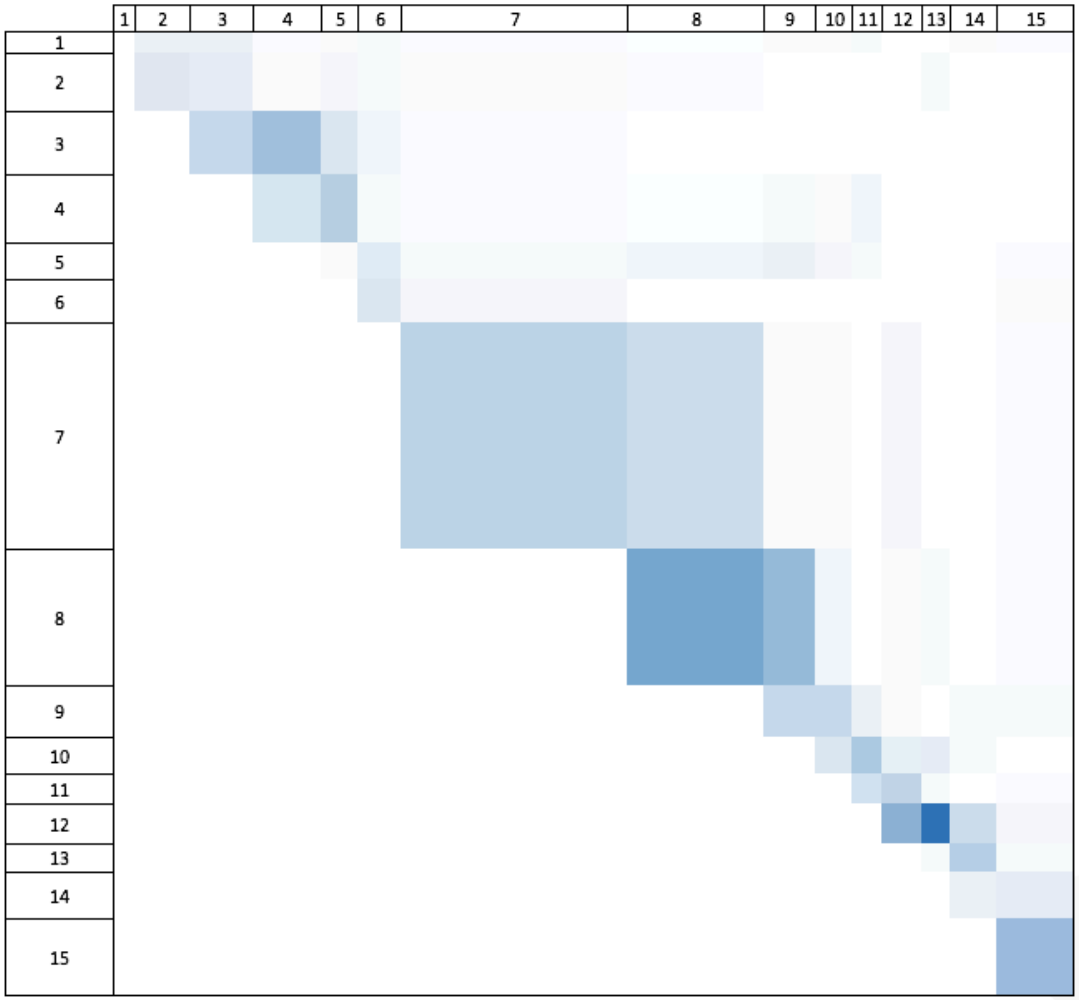
Results (I)

- **Backward jumps heatmap.**



Results (II)

- **Forward jumps heatmap.**



Conclusions and future research

- Heatmaps based on storyboard can be used to detect problematic scenes.
- Heatmaps show that majority of participants jumps within the same scene and follow the linearity of time.
- Can be used to analyze bottlenecks and hot spots.
- Automatic storyboard extraction.
- Improve boudaries between consecutive scenes

Thank you!

jtriaypa[at]uoc[dot]edu

@jtriayp