Video Analytics

INFO 4100 Learning Analytics

[[Shuhui Zhu, sz649]]

In this homework, you will learn how to analyze video interaction data to generate video analytics.

Learning Objectives:

1. Exploring the structure of video interaction data
2. Identifying parts of the video with increased activity
3. Deciding what video analytics to report back to learners and instructors

You are given timestamped video interaction events for this video from an online course: <https://youtu.be/qKNb8YQYTZg>.

Dataset columns:

* id is a hashed learner id
* time is the exact time of the event
* order counts observed events in temporal for each id
* seconds measures time between events for the same person
* event\_type
* video\_current\_time is the time in the video the event occured
* video\_new/old\_speed is for tracking speed\_change events
* video\_old/new\_time is for tracking seek events

# Part 1: Explore video data

**Question 1:** What event types are in the dataset and how many of each?

#######################################  
####### BEGIN INPUT: Question 1 #######  
#######################################  
table(vid$event\_type)

##   
## load\_video pause\_video play\_video seek\_video   
## 3060 7572 12502 8008   
## speed\_change\_video stop\_video   
## 172 587

#######################################  
#######################################

**Question 2:** How many users watched the video at all? How many video events do users have on average?

#######################################  
####### BEGIN INPUT: Question 2 #######  
#######################################  
  
sub\_v <- vid[which(vid$event\_type == 'play\_video'), ]  
length(unique(sub\_v$id)) # 934 users watched the video

## [1] 934

sub\_avg <- vid %>%  
 group\_by(id) %>%  
 summarise(cnt = n())  
  
mean(sub\_avg$cnt) # 20.66 video events on average

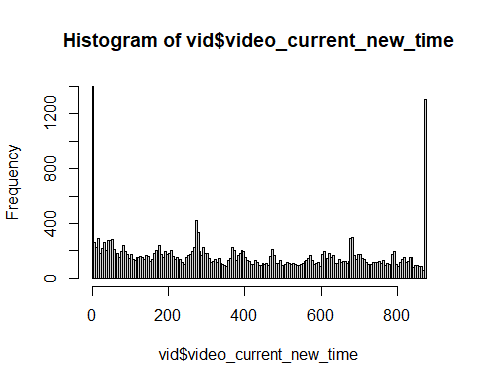
## [1] 20.66127

#######################################  
#######################################

**Question 3:** Using a histogram, plot how much activity occurred throughout the video in total (x-axis = time in video; y-axis = event count). Tip: Set breaks=300 in the hist() function for better resolution. Note that seek\_video events have the video\_current\_time field missing so you should use video\_old\_time instead.

You should see a spike at the start, the end, one at around 300, and another around 700.

#######################################  
####### BEGIN INPUT: Question 3 #######  
#######################################  
  
vid$video\_current\_new\_time <- vid$video\_current\_time  
  
vid$video\_current\_new\_time[is.na(vid$video\_current\_time)] <- vid$video\_old\_time[is.na(vid$video\_current\_time)]  
  
hist(vid$video\_current\_new\_time, breaks=300)



######################################  
#######################################

**Question 4:** At what point in the video do many learners pause? Using a precision of 10 seconds (meaning that you should round numbers to the nearest 10 seconds, so e.g. 123 becomes 120 and 47 becomes 40), report the 3 most common times in seconds. Then go to [the video lecture](https://www.youtube.com/watch?v=qKNb8YQYTZg&t=870s) and look up what happens at those times (note that you can type in the seconds in this URL so you don’t need to convert it). Say why you think learners paused there.

A: Top three times where learners paused are at 870, 270, and 0. Pausing at 870 could be since its near the end, students pause to end early. Pausing at 270 could be the instructor initiated a run command, which may take some time for students to either figure out and wait for results. 0 is around the start, students may not be prepared with the class yet. They need to get the program running and other preparation tasks.

#######################################  
####### BEGIN INPUT: Question 4 #######  
#######################################  
## round the time first!  
library(plyr)

## Warning: package 'plyr' was built under R version 4.0.5

## ------------------------------------------------------------------------------

## You have loaded plyr after dplyr - this is likely to cause problems.  
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:  
## library(plyr); library(dplyr)

## ------------------------------------------------------------------------------

##   
## Attaching package: 'plyr'

## The following objects are masked from 'package:dplyr':  
##   
## arrange, count, desc, failwith, id, mutate, rename, summarise,  
## summarize

## The following object is masked from 'package:purrr':  
##   
## compact

sub\_v\_pause <- vid[which(vid$event\_type == 'pause\_video'), ]  
sub\_v\_pause$video\_current\_time\_rnd <- round\_any(sub\_v\_pause$video\_current\_time, 10, f = floor)  
freq\_c <- table(sub\_v\_pause$video\_current\_time\_rnd)  
# freq\_c[order(freq\_c[,1])]  
freq\_c <- as.data.frame(freq\_c)  
head(freq\_c[order(freq\_c[,2], decreasing = TRUE),], 10)

## Var1 Freq  
## 88 870 592  
## 28 270 252  
## 1 0 186  
## 2 10 173  
## 5 40 162  
## 6 50 143  
## 4 30 139  
## 69 680 134  
## 30 290 131  
## 27 260 128

#######################################  
#######################################

**Question 5:** At what point in the video do many learners seek forwards and when do they seek backwards in the video (ignore for now where they seek to)? Using a precision of 10 seconds, report the 3 most common times in seconds for moving back, 3 most common for moving forward. Then go to [the video lecture](https://www.youtube.com/watch?v=qKNb8YQYTZg&t=190s) and look up what happens at those times. Why might learners move? Forward: 0, 170, 180 0 - Many students may watch the video more than once, students will jump to other parts from the beginning. 170 - I find this part the pace is a bit slow; learners may find this part easy and a good idea to skip  
180 - Similar to 170, it’s more about checking the result; students may feel bored in listening the interpretation of the results

Backward: 280, 270, 190 280 - typing involved, more actions in the next 10 seconds, it takes time to process both the typing and listening 270 - speaking speed and execution are quite fast, learners may not follow the instruction 190 - typing involved, speaking speed is fast, learners may not catch the explanation

#######################################  
####### BEGIN INPUT: Question 5 #######  
#######################################  
  
sub\_v\_seek <- vid[which(vid$event\_type == 'seek\_video'), ]  
sub\_v\_seek$diff\_back\_forward <- sub\_v\_seek$video\_new\_time - sub\_v\_seek$video\_old\_time   
  
sub\_v\_seek <- sub\_v\_seek %>% mutate(back\_forward = ifelse(diff\_back\_forward >= 0 ,'forward', 'back'))  
sub\_v\_seek\_forward <- sub\_v\_seek[which(sub\_v\_seek$back\_forward == 'forward'),]  
sub\_v\_seek\_back <- sub\_v\_seek[which(sub\_v\_seek$back\_forward == 'back'),]  
  
# mv foward  
sub\_v\_seek\_forward$video\_current\_time\_rnd <- round\_any(sub\_v\_seek\_forward$video\_current\_new\_time, 10, f = floor)  
freq\_c\_fwd <- table(sub\_v\_seek\_forward$video\_current\_time\_rnd)  
# freq\_c[order(freq\_c[,1])]  
freq\_c\_fwd <- as.data.frame(freq\_c\_fwd)  
head(freq\_c\_fwd[order(freq\_c\_fwd[,2], decreasing = TRUE),],10)

## Var1 Freq  
## 1 0 215  
## 18 170 84  
## 19 180 84  
## 29 280 80  
## 5 40 78  
## 15 140 78  
## 38 370 74  
## 7 60 72  
## 8 70 69  
## 2 10 68

# mv back  
sub\_v\_seek\_back$video\_current\_time\_rnd <- round\_any(sub\_v\_seek\_back$video\_current\_new\_time, 10, f = floor)  
freq\_c\_back <- table(sub\_v\_seek\_back$video\_current\_time\_rnd)  
# freq\_c[order(freq\_c[,1])]  
freq\_c\_back <- as.data.frame(freq\_c\_back)  
head(freq\_c\_back[order(freq\_c\_back[,2], decreasing = TRUE),],10)

## Var1 Freq  
## 29 280 122  
## 28 270 112  
## 20 190 110  
## 6 50 93  
## 69 680 86  
## 70 690 82  
## 9 80 76  
## 39 380 74  
## 88 870 73  
## 85 840 67

#######################################  
#######################################

**Question 6:** At what point in the video do many learners seek backwards in the video and where do they go? Using a precision of 10 seconds, report the 3 most common pairs of times in seconds like this <from, to>. Then go to [the video lecture](https://www.youtube.com/watch?v=qKNb8YQYTZg&t=270s) and look up what happens at those times. Why do you think learners moved there?

The top three are 280 to 270; 270 to 260; 680 to 670. Between these three time slots are quick executions of code. The content on the screen shown are different after execution. Students may need to refer to the code.

#######################################  
####### BEGIN INPUT: Question 6 #######  
#######################################  
sub\_v\_seek\_back$video\_old\_time\_rnd <- round\_any(sub\_v\_seek\_back$video\_old\_time, 10, f = floor)  
sub\_v\_seek\_back$video\_new\_time\_rnd <- round\_any(sub\_v\_seek\_back$video\_new\_time, 10, f = floor)  
sub\_v\_seek\_back$concat\_old\_new\_time\_rnd <- paste(sub\_v\_seek\_back$video\_old\_time\_rnd, sub\_v\_seek\_back$video\_new\_time\_rnd, sep = ' ' )  
  
freq\_c\_back\_from\_to <- table(sub\_v\_seek\_back$concat\_old\_new\_time\_rnd)  
# freq\_c[order(freq\_c[,1])]  
freq\_c\_back\_from\_to <- as.data.frame(freq\_c\_back\_from\_to)  
head(freq\_c\_back\_from\_to[order(freq\_c\_back\_from\_to[,2], decreasing = TRUE),],10)

## Var1 Freq  
## 206 280 270 49  
## 191 270 260 42  
## 670 680 670 38  
## 463 50 40 37  
## 192 270 270 34  
## 584 610 600 33  
## 95 190 180 32  
## 94 190 170 30  
## 314 370 360 30  
## 2 10 0 27

#######################################  
#######################################

**Question 7:** Are students more likely to speed the video up or slow it down? Report the proportion speed-ups relative to all speed changes. How do you interpret this? Students are more likely to speed up. The proportion is 0.575. There are more students who adjust the speed faster, whereas there are also nearly half of the students who adjust the speed slower.

#######################################  
####### BEGIN INPUT: Question 7 #######  
#######################################  
  
vid$video\_speed\_diff <- vid$video\_new\_speed - vid$video\_old\_speed  
vid <- vid %>% mutate(fast\_slow = ifelse(video\_speed\_diff >= 0 ,'fast', 'slow'))  
table(vid$fast\_slow)

##   
## fast slow   
## 99 73

# based on result from the table  
99/(99+73)

## [1] 0.5755814

#######################################  
#######################################

**Question 8:** Based on the video analytics you have done (or any additional analyses you’d like to do), what information would you give (a) a learner about to watch this lecture video, and (b) the instructor of the lecture video? Make emprically grounded recommendations for each stakeholder.

* **Recommendation to learners:** Try to be prepared before opening a video. Otherwise, you will interrupt the watching experience. There are other learners who adjust the speed for either faster or slower. It is fine for you to adjust a speed that you find comfortable. There are several points in the video many other students find hard to follow, like at 270, 280, 190 second.
* **Recommendation to instructor:** Instructors should keep the code and result output appear at the same time on the screen. Students may fall behind in typing code and may need to refer to code to understand the result.

# Self-reflection

**Briefly summarize your experience on this homework. What was easy, what was hard, what did you learn?**

* This homework was easier to me compared to last homework. I think this analysis is very interesting and for some online course, video editors could play around with the finding of the result to make the viewing experience even better.

# Submit Homework

This is the end of the homework. Please **Knit to Word**. The resulting file has to show both the R code and R output. Upload it on the EdX platform before the due date.