CS539 –Advanced Topics on Wireless Networks and Mobile Computing

Department of Computer Science
University of Crete

1st Assignment: Analysis of user experience

In this assignment you will take a glance at some of the issues that occur during video streaming, like in the case of YouTube streaming. Probably all of have experienced during watching a video on YouTube rebufferings (screen freezing/loading), positive or negative bitrate changes that result in improvement or making poorer the video resolutions, startup delay, advertisements. All these data along with the duration of the viewed video, the resolution level at each time instance, can be collected in order to evaluate the impact they have on the user engagement to the service.

Here you are asked to analyze the impact of some of these features for thousands of recordings collected from real YouTube views all over the world.

The data to analyze include for each session (i.e., viewing episode) the following attributes (the numbers here correspond to columns of the data provided):

1. number of rebuffering events (RBs)
2. total duration of RBs in sec
3. mean weighted bitrate in Kbps (mwb)
4. number of positive bitrate changes (BR+)
5. number of negative bitrate changes (BR-)
6. initial buffering time (or startup delay)
7. the percentage of the video that the user has watched (vwp)
8. abandonment status: 0 (session completed), 1 (stopped due to buffering stalling), 2 (stopped by user)
9. video duration
10. session duration

Use MATLAB to answer the following questions:

1. Calculate the min, mean, median and max value, as well as the standard deviation of the attributes provided (except from the abandonment type). Create a corresponding table in your report to present the results.

2. Find the mean RB duration of the RBs recorded. Save these values in a vector called mean_rb_dur. If a session does not have RBs, the corresponding value should be 0. Plot the distribution of this vector.

3. **a.** Observe the attribute distributions and make comments about them. Provide the ECDFs.

b. What is the dominant value of each variable and with what frequency does it occur? Dominant value is the value that appears more often in the dataset. (You can use a combination of *histogram* and *unique* MATLAB functions). Provide corresponding plots.

4. **a.** Is there any linear correlation between number of RBs and RB duration for session in which more than half of the video has been watched? Justify your

answer. You can show this with use of scatter (graphic way) and corr (statistical way) MATLAB functions for this specific question.

5. Plot and compare the distributions at each of the following cases with use of Smirnov-Kolmogorov test to examine whether or not there is a statistically significant difference between them. Justify your answers:

a: video watching percentage of sessions with BR+ *vs.* the video watching percentage of sessions with BR-.

(Here the BR changes should be the only impairments, so no RBs should occur.)

b: vwp of abandoned sessions (type 1 and 2) with BR- but without RB *vs.* vwp of abandoned (type 1 and 2) sessions without RB nor BR.

6. The user engagement can be quantified with the percentage of video content the user has watched until the end of the session (vwp). Use LASSO regression to find which of the session characteristics have high predictive power on user engagement. Use attributes 1-6 and 8-9 and/or create any new if you think this would help. Explain your answer.

Deliverables:
-Report (.pdf file) that includes all the results/plots/comments needed.
-MATLAB (or R or Python) code you wrote

It is good for each question to print corresponding messages with your code, and for the plots to have proper title and axes' names so that it is easier for the reader of your code/report to understand what is presented.
Moreover, when you are studying ECDF plots sometimes it's useful to limit the axes properly to make the plot more readable (especially in the case of extremely high values and long tails). In such cases, limit properly the x axis of the plot the 95^{th} percentile (to do so use xlim function in MATLAB).