Unska 3, 10000 Zagreb, Croatia

SUPLLEMENTARY MATERIAL FOR DATASET

Subject: **This document is supplementary material for a dataset published in conjunction with the paper "Impact of User Playback Interactions on In-Network Estimation of Video Streaming Performance" by Ivan Bartolec, Irena Orsolic, and Lea Skorin-Kapov.**

The dataset is part of the paper "**Impact of User Playback Interactions on In-Network Estimation of Video Streaming Performance**" published in IEEE Transactions on Network and Service Management (TNSM) journal with DOI **10.1109/TNSM.2022.3180114**. If you use this dataset in any way, please make sure to reference the previously mentioned paper.

The dataset is comprised of four CSV files, one for each of the playback interactions that were triggered while the video was being played. The dataset contains 7424 videos in total. Among these videos, 2030 had no interactions (dataset NoInteraction), 1749 had a single pause (dataset Pause), 1808 had a single seek during video playback (dataset Seek), and 1837 were abandoned prior to completion (dataset Abandon). Each file includes both application data (information downloaded from the mobile application) and network traffic statistics (derived from traffic and used as features in machine learning training). The data for each second of the video session (one row) contains all application layer data and network traffic statistics for that second or aggregated values from the beginning of the video session to that second of video playback.

APPLICATION DATA

videoID	video identifier
interval	timestamp
interaction	interaction type (null/pause/seek)
bitrate	video bitrate in kbps
resolution	video resolution (144/240/360/480/720/1080/1440)
buffer	miliseconds of video stored in the buffer
frame	number of frames dropped and the number of displayed frames
bitrate2	video bitrate (low/high)
resolution3	video resolution (low/medium/high)
resolution2	video resolution (sd/hd)
stall1	video stalling (n/y)
stall2	video stalling occurred due to seek interaction and initial delay (n/y)
interaction_occured	has interaction occurred (n/y)

NETWORK TRAFFIC STATISTICS (FEATURES)

All features come in downlink traffic (dl) and uplink traffic (ul) variants. wN denotes window-based features, where N can be 1,2,3,5,10, or 20. The notation for slotted features is slN, where N is either 1,5, or 10. Statistical methods such as mean, median, minimum, maximum, standard deviation, skewness, kurtosis, and variance are used to derive each feature.

pckt_count_wN	packet count (in the last N seconds)
pckt_count_gt100	packet count where all the packets smaller than 100B are ignored (in the last N seconds)
active_time_wN	percentage of the window that is used for transmission (in the last N seconds)
pckt_size_wN	packet size (in the last N seconds)
pckt_size_gt100_wN	packet size where all packets smaller than 100B are ignored (in the last N seconds)
iat_wN	interarrival time of packets (in the last N seconds)
throughput_wN	throughput (in the last N seconds)
slN_pckt_size	slotted packet size
slN_pckt_size_gt100	slotted packet size where all the packets smaller than 100B are ignored
slN_throughput	slotted throughput
distribution_iat	interarrival time of packets from video start to this exact timestamp
distribution_pckt_num	packet count from video start to this exact timestamp
distribution_pckt_num_gt100	packet count from video start to this exact timestamp where all packets smaller than 100B are ignored

RELEVANT FEATURES FOR MACHINE LEARNING MODELS

SESSION-BASED CLASSIFICATION MODELS

Average video bitrate in 2 classes

'sl5_variance_dl_pckt_size_gt100',
'sl1_min_ul_pckt_size',
'sl5_mean_ul_pckt_size',
'sl10_stdev_ul_pckt_size',
'sl5_min_dl_throughput',
'sl10_variance_ul_throughput',
'min_distribution_dl_iat',
'variance_distribution_ul_iat',

'min_distribution_ul_iat',
'min_distribution_ul_pckt_num'

Averge video bitrate in 3 classes

'sl5_min_dl_pckt_size',
'sl1_mean_dl_pckt_size_gt100',
'sl5_min_dl_pckt_size_gt100',
'sl1_min_ul_pckt_size',
'sl10_variance_ul_pckt_size',
'sl10_stdev_ul_pckt_size',
'sl5_min_dl_throughput',
'sl5_min_ul_throughput',
'min_distribution_dl_pckt_num',
'median_distribution_ul_pckt_num_gt100'

Longest played resolution in 3 classes

'sl1_mean_dl_pckt_size',
'sl10_median_dl_pckt_size',
'sl5_skewness_dl_pckt_size_gt100',
'sl1_min_ul_pckt_size_gt100',
'sl1_min_dl_throughput',
'sl5_min_dl_throughput',
'sl1_min_ul_throughput',
'min_distribution_dl_iat',
'min_distribution_ul_pckt_num',
'min_distribution_ul_pckt_num_gt100'

Longest played resolution in 6 classes

sl1_kurtosis_dl_pckt_size',
'sl5_kurtosis_dl_pckt_size',
'sl1_max_dl_pckt_size_gt100',
'sl1_min_dl_pckt_size_gt100',
'sl5_variance_ul_pckt_size',
'sl10_min_ul_pckt_size',
'sl1_min_ul_pckt_size_gt100',
'variance_distribution_dl_iat',
'min_distribution_dl_iat',
'min_distribution_ul_pckt_num_gt100'

Stalling occurred in 2 classes

'sl1_variance_dl_pckt_size',
'sl10_min_ul_pckt_size_gt100',
'sl1_stdev_dl_throughput',
'sl1_min_dl_throughput',
'sl10_variance_dl_throughput',
'sl10_stdev_dl_throughput',
'max_distribution_dl_iat',
'max_distribution_ul_pckt_num',
'max_distribution_dl_pckt_num_gt100',
'variance_distribution_ul_pckt_num_gt100'

REAL-TIME CLASSIFICATION MODELS

Bitrate in 2 classes

'sl10_max_dl_throughput'
'max_distribution_dl_pckt_num_gt100'
'sl1_mean_dl_pckt_size'
'sl5_mean_dl_pckt_size'
'sl5_median_dl_pckt_size'
'sl10_median_dl_throughput'
'sl1_variance_dl_throughput'
'sl1_max_dl_throughput'
'sl5_stdev_dl_throughput'
'sl10_max_ul_throughput'

Bitrate in 3 classes

'max_distribution_dl_pckt_num_gt100'
'sl1_max_dl_throughput'
'variance_distribution_dl_pckt_num'
'sl10_max_dl_throughput'
'sl10_max_dl_pckt_size_gt100'
'sl1_variance_dl_throughput'
'sl5_stdev_dl_throughput'
'sl5_max_dl_throughput'
'sl10_median_dl_pckt_size_gt100'
'mean_distribution_dl_pckt_num'

Resolution in 3 classes

'max_distribution_dl_pckt_num_gt100'
'sl1_max_dl_throughput'
'variance_distribution_dl_pckt_num'
'sl10_max_dl_throughput'
'sl10_max_dl_pckt_size_gt100'
'sl1_variance_dl_throughput'
'sl5_stdev_dl_throughput'
'sl5_max_dl_throughput'
'sl10_median_dl_pckt_size_gt100'
'mean_distribution_dl_pckt_num'

Resolution in 6 classes

'max_distribution_dl_pckt_num_gt100'
'sl10_max_dl_throughput'
'sl10_median_dl_pckt_size_gt100'
'variance_distribution_dl_pckt_num'
'sl5_mean_dl_pckt_size'
'sl5_median_dl_pckt_size'
'sl1_variance_dl_throughput'
'sl1_max_dl_throughput'
'sl5_stdev_dl_throughput'
'sl5_max_dl_throughput'

Stalling occurred in 2 classes

'max_distribution_dl_pckt_num_gt100'
'sl1_max_dl_throughput'

'median_distribution_ul_iat'
'sl5_median_dl_pckt_size'
'mean_distribution_ul_iat'
'sl1_variance_dl_throughput'
'ul_median_iat_w2'
'ul_median_iat_w20'
'sl10_median_dl_pckt_size_gt100'
'sl10_stdev_ul_throughput'