Online Video Characteristics and Transcoding Time Dataset Data Set

Download: Data Folder, Data Set Description

Abstract: The dataset contains a million randomly sampled video instances listing 10 fundamental video characteristics along with the YouTube video ID.

Data Set Characteristics:	Multivariate	**Number of Instances:**	168286	**Area:**	Computer
Attribute Characteristics:	Integer, Real	**Number of Attributes:**	11	**Date Donated**	2015-05-19
Associated Tasks:	Regression	**Missing Values?**	N/A	**Number of Web Hits:**	29184

Source:

Tewodros Deneke, tdeneke '@' abo.fi

Data Set Information:

The presented dataset is composed of two tsv files named 'youtube_videos.tsv' and 'transcoding_mesurment.tsv'. The first contains 10 columns of fundamental video characteristics for 1.6 million youtube videos; It contains YouTube video id, duration, bitrate(total in Kbits), bitrate(video bitrate in Kbits), height(in pixle), width(in pixles), framrate, estimated framerate, codec, category, and direct video link. This dataset can be used to gain insight in characteristics of consumer videos found on UGC(Youtube).

The second file of our dataset contains 20 columns(see column names for names) which include input and output video characteristics along with their transcoding time and memory resource requirements while transcoding videos to diffrent but valid formats. The second dataset was collected based on experiments on an Intel i7-3720QM CPU through randomly picking two rows from the first dataset and using these as input and output parameters of a video transcoding application, ffmpeg 4 . In section 6 we will use the second dataset to build a transcoding time prediction model and show the significance of our datasets.

Attribute Information:

id = Youtube videp id
duration = duration of video

bitrate bitrate(video) = video bitrate
height = height of video in pixles
width = width of video in pixles
frame rate = actual video frame rate
frame rate(est.) = estimated video frame rate
codec = coding standard used for the video
category = YouTube video category
url = direct link to video (has expiration date)
i = number of i frames in the video
p = number of p frames in the video
b = number of b frames in the video
frames = number of frames in video
i_size = total size in byte of i videos
p_size = total size in byte of p videos
b_size = total size in byte of b videos
size = total size of video
o_codec = output codec used for transcoding
o_bitrate = output bitrate used for transcoding
o_framerate = output framerate used for transcoding
o_width = output width in pixel used for transcoding
o_height = output height used in pixel for transcoding
umem = total codec allocated memory for transcoding
utime = total transcoding time for transcoding

Relevant Papers:

@INPROCEEDINGS{6890256,
author={Deneke, T. and Haile, H. and Lafond, S. and Lilius, J.},
booktitle={Multimedia and Expo (ICME), 2014 IEEE International Conference on},
title={Video transcoding time prediction for proactive load balancing},
year={2014},
month={July},
pages={1-6},
keywords={prediction theory;resource allocation;transcoding;video coding;video streaming;input video stream;proactive load balancing;video transcoding time prediction;Bit rate;Codecs;Load management;Load modeling;Predictive models;Transcoding;YouTube;Load Balancing;Machine Learning;Prediction;Transcoding},
doi={10.1109/ICME.2014.6890256},}

Citation Request:

If you have no special citation requests, please leave this field blank.