**3.4.2**

**Identify the predominant Ad topics**

**1. Idea:**

Detect the ads by feature extraction of video.

**Motivation**

There is an increased amount of video and audio content broadcast and streamed everywhere today, specifically on websites. Such content needs to be frequently analyzed for a variety of reasons and applications – such as searching, indexing, summarizing, etc. One general area is identifying content of frames or even a number of frames altogether such as (gambling, prescription drugs, CBD drugs, easy-money remote jobs and Ponzi schemes)

**Problem Statement**

Design an algorithm to automatically identify advertisements from the video (and corresponding audio) which is interspersed with advertisements in websites. Furthermore, design a process based on specific images to detect the type of the content in the video.

**Keywords**

Frame: a single still image from a video, eg NTSC - 30 frames/second, film – 24frames/second

Shot: sequence of frames recorded in a single camera operation

Sequence or Scenes: collection of shots forming a semantic unit which conceptually may be shot at a single time and place

**Process**

1. Read in the input video/audio – remember we might not be able to fit the entire content in memory for processing.
2. Break the input video or audio into a list of logical segments – shots. We will by extracting relevant features, build a learning model and remove the advt. segments in the test data(audio segments) using Machine Learning techniques. Extracted critical features from radio recordings to classify the input samples into songs and advertisements using the Support Vector Machines.

* Train the SVM classifier with 1000 samples to classify test samples into two classes – Ad or Song.
* Wrote a correction algorithm to improve accuracy.

1. Give each shot a variety of quantitative weights such as – length of shot, motion characteristics in the shot, audio levels, color statistics etc.
2. Using the above characteristics, decide whether a shot or a group of adjacent shots might be an advertisement
3. If brands (gambling, prescription drugs, CBD drugs, easy-money remote jobs and Ponzi schemes) are detected, collect the data and train a model based on this data.

**2. Idea:**

All what we need are the filter rules which can be applied on the URLs of HTML elements to check if they are ads or not. **Note:** We don't really need to render the elements in order to use Adblock.

If we detected correctly:

It would be nice to train regular expressions using input webpages for each advertisement.