Detecting ads from captions

Vishnu Lokhande

April 22$^{\text{nd}}$, 2017

1 Objective

The goal is to detect whether a video segment is an advertisement or not based on the captions extracted from the video segments.

2 Data description

Different video streams of length ranging from 30 to 40 minutes is provided. Each such video stream is divided into 60 sec video segments. Each segment has about 25 captions encoded into it. This amounts to a single caption for every 2.4 second of video.

3 Caption extraction from video segments

Captions can be extracted from the video segments using *ffmpeg* software. A couple of observations while extracting the Ads are as follows

- If the 60 sec segment is entirely contained of ADs, then that video segment does NOT have any captions encoded with it.
- If the 60 sec segment is mainly an AD. That is, if nearly three-fourth of the segment is an AD, then we do NOT have captions for this section of the video. Only the non-AD content of the segment has captions encoded.
- If the 60 sec segment is partly an AD, then, we have caption for the entire 60 seconds of the video.

4 Rule based algorithm for Ad detection

It has been observed that the captions that can be extracted from the Ad portion of the video segment is in all small letters and the caption for non-Ad portion is in all capital letter. A simple algorithm to detect the case of the letters in the caption can be used to detect the Ads.

5 Learning based algorithm for Ad detection

This approach has been attempted considering all the captions to be in lowercase. The different stages of the approach are explained as follows

5.1 Feature Extraction

5.1.1 Pre-processing

Stop-words such as 'the','a','to', etc., have been removed from the captions.

5.1.2 Bag-of-words

Using the data from the entire video stream, a dictionary of words has been prepared containing all the distinct words from all the captions of the video stream. Each caption is then converted into a vector of 1s and 0s of dimension equal to the dimension of the dictionary. A '1' is assigned to an element in the vector if the corresponding word of the caption is present in the dictionary and vice-versa.

5.1.3 TF-IDF measures

The vector which is assigned to a caption is then converted into TF-IDF measures.

5.2 Classification

Experiments were conducted on FOX news video stream. This video stream was also labelled manually in order to perform supervised learning.
Three classifiers were tested which are Multinomial naive bayes, stochastic gradient descent classifier and bernoulli naive bayes. Performance was measured using thee-fold cross-validation. It was observed that SGD classifier was performing better than the other two.

5.3 Results

The data is skewed, that is, there are less number of captions for the Ads than the non-Ads. Thus, in this case, precision and recall measure is more helpful than accuracy measure.
The average precision for the captions from Ads over three-fold cross validation is 79% and average recall is 43%. This score is for SGD classifier. These scores show that we have a satisfactory classification model.