**International Journal of Multimedia Data Engineering and Management**
Volume 10 • Issue 2 • April-June 2019

# A Framework Model for Integrating Social Media, the Web, and Proprietary Services Into YouTube Video Classification Process

Mohamad Hammam Alsafrjalani, University of Miami, Miami, USA

## ABSTRACT

Online video streaming has gained ubiquity in disparate educational, governmental, and corporate environments. The ubiquity of these videos elicits new challenges in video classification that is used to promote relative videos and block unwanted content. These challenges include the incorporation of contextual information, rapid development of the ad-hoc query modules, and keeping pace with contemporary contextual information. In this article, the authors present a framework model for incorporating contextual information into video classification information. To illustrate the model, the authors propose a framework which comprises video classification, web search engines, social media platforms, and third-party classification modules. The modules enable the framework's flexibility and adaptivity to different contextual environment—educational, governmental, and corporate. Additionally, the model emphasizes standardized module interface to enable the framework's extensibility and rapid development of future modules.

## KEYWORDS

Ad-Hoc Modules, Cybersecurity, Social Media, Twitter, Videos Classification, YouTube

## 1. INTRODUCTION

The importance of video classification has increased in the last decade due to the ubiquity and exponential growth of crowd-generated videos—more than 80% of the internet traffic will be for video streaming (Cisco, 2017). Demands for classifying videos are elicited by various corporate, government, and educational policies that require identification of harmful videos (e.g., phishing and spam), hate- and crime-promoting content, pornography, and/or cyber bullying, which can be spread through online video streaming platforms. Alternatively, the classification information can be used to promote relative content, generate targeted advertisement, and retain more users (Duverger & Steffes, 2012; Xu, Zhang, et al., 2008).

To classify YouTube videos, machine learning algorithms (e.g. Convolutional Neural Networks (CNN) (Karpathy et al., 2014)) are used to extract key features from the videos' frames (e.g., (Roach, Mason, & Pawlewski, 2001)), text (e.g., (Brezeale & Cook, 2006)), audio (e.g., (Z. Liu, Wang, & Chen, 1998)), or a combination of these data (e.g., (Qi, Gu, Jiang, Chen, & Zhang, 2000)). This

DOI: 10.4018/IJMDEM.2019040102

information is used to train the models offline using a set of pre-downloaded videos. After the models are trained, they are used to predict the class of new, unknown videos. However, the drawbacks of these approaches include lengthy training time and high computation complexity, and a priori access to training videos.

To reduce time and computation complexity, other research targeted key frames and segments (Lu, Drew, & Au, 2001), the video's texts (Huang, Fu, & Chen, 2010), such as the title, description, and/or comments. However, these approaches can only predict classification of new videos using training on pre-determined features that were available during training time, and they did not consider new, unprecedented classes (e.g., a new online challenge). Furthermore, the exponential growth of the number of video creators on social media—over 1 million YouTube channels (YouTube, 2017) and 8 billion Facebook videos watched per day (Tang et al., 2017), the number of new, subjective video classes will increase beyond what a single model can determine. Furthermore, the subjectivity of the information technology's (IT) policies deployed at different entities—government agencies, corporate office, and education institution, exacerbate the relativity of the classification information. Since these policies are focused on the entity's demands, the policies require adaptivity to the contextual environment in which the video is being streamed.

A generic video classification information (sport, news, entertainment, …) must be accompanied with contextual information. Whereas the generic classification information provides valid information, awareness of the contextual environment will impact the IT decision about these videos. For instance, a video that was classified into commercial advertisement, as a result of pre-trained classification model, could be also classified as an online challenge based on recent events in the online and social media. Since the recent events, and possibly the category itself, were unavailable during the module training period, the trained module would fail to predict the online challenge as a category of the video. If the recent events, contextual information, is obtained as part of a classification framework, the IT policies can provide more accurate decisions.

To keep up with the correct usage of the video classification information, the contextual information must be captured in relatively short time by querying information available on platforms other than the video streaming platforms—such as news platforms, law and legislation bills, technical specifications, etc. Furthermore, since not all the platforms are always relative to query, querying the information should be an ad-hoc process that provides only the necessary information for the decision-making policy. Based on this information we deduce that the next challenges in video classification include not only video analysis, but also the contextual environment in which the classification is being used. Incorporation of contextual information, rapid development of the ad-hoc query modules, and keeping pace with contemporary contextual information must be addressed to meet future challenges in video classification.

In this paper we propose a framework model that enables rapid development of modules that perform contextual-information retrieval in conjunction with video classification information. We illustrate our model by developing a modular framework. The framework's modules obtain classification information from the video streaming platform and contextual information from social media, web content and third-party domain classification services. To enable flexibility, the modules work independently to make fine-grained, per-module classifications and collaboratively to make coarse-grained, final classifications decisions. Additionally, to enable rapid adjustment to disparate environment, the modules can be used in ad-hoc fashion. Furthermore, to enable the framework's extensibility, the modules have a standardized interface and communication protocols so that future modules can be easily integrated into the framework.

To evaluate our proposed framework, we classified 25 randomly selected YouTube videos and tested the accuracy of these classifications against human verification. Furthermore, we evaluate the framework modules in educational, governmental, and corporate environment and calculated the classification accuracies under different classification requirements. Our results revealed that our framework classified the videos with up to 90% accuracy, and classifications process completed