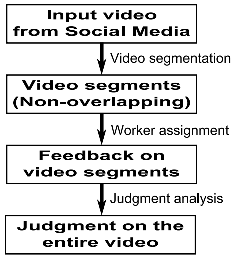

Figure 1: Flowchart of the proposed model.

judge whether it is unsafe). Each segment is judged by a set of workers represented as $W_i = \{w_{i1}, w_{i2}, \ldots, w_{im_i}\}$, for all i, and m_i denotes the number of workers for the i^{th} segment. Note that, the set of workers for each segment are not necessarily distinct. The workers for each segment provide their opinions in the form of Yes (Y) or No (N), which signify whether the video segment is accepted or rejected as being safe, respectively. After receiving the responses on each segment, decision is taken by majority voting. So, for each segment we get a judgment either as Yes or No. The video will be considered safe (suitable) if all of segments are judged as Yes, otherwise not. A schematic view of the entire approach is shown in Fig. 1 and is detailed hereunder.

Segmentation of videos: The foremost step is to decompose a video into minimum time slots so that it contains sufficient amount of information for making a judgment. We adopt a recent approach of Deza et al. explaining the way to improve the human performance in a realistic sequential visual search task (**?**). We consider the (minimum) time duration τ required for understanding the video segment by the workers.

Assigning crowd workers: Thereafter comes the proper selection of workers for assigning the video segments. If a segment is judged by a single worker then the outcome is not always reliable, so we assign a set of workers to solve a common segment independently. Online viewers are of two types – signed (mobile users and others who enter into the platform with unique identity) and unsigned (whose profile is unknown). Primarily, we prefer the signed viewers as the crow workers to judge the segments. If signed workers are not sufficient, then we will consider the unsigned workers. Based on the performance, we are awarding the workers with credit points to motivate them. We also track the workers' profile for a better assignment. If a video is posted with American accent then it is obvious that it will be perfectly perceived by a worker who is native American.

Judgment analysis on each segment: As the videos are coming in a streamline, the accuracy of crowd workers are computed with respect to a fixed sized sliding window. Majority voting is initially applied to a particular video and based on the agreement of worker's opinion with the majority, the accuracy of a particular crowd worker is calculated. These accuracy values are then incorporated in the final stage while aggregating the opinions. As it is obvious that accuracy of a worker is not constant for different time instants, therefore the accuracy scores are tracked after certain time intervals. Thus, majority voting is repeatedly applied for a particular video within particular time interval. The other issue that we consider is the biasness of workers over a particular video. Biasness of crowd workers means while providing their opinions they might be inclined to either Yes or No. To identify these characteristics, some videos (as test cases) having ground truth label are included and based on the responses on them the biasness is determined.

Final decision making: Final decision is taken based on the collective judgment on all the segments. If all the segments are marked Yes then it means that the entire video is acceptable by the workers. So, the video is safe for the platform. If any of the segment is rejected by the assigned group, then the video is unsafe.

Challenges

Every second massive amount videos get uploaded by the different users. Judging a streaming video in real-time by the crowd workers is a demanding job. It depends on many factors such as the selection of workers, fixing the number of workers for each segment (**?**), response time of each worker, duration of each video segment, etc. It is also challenging to find the ways to segregate a video into overlapping segments for reducing the misinterpretation at the time of judgement. Attributes of a video and demography of the assigned workers play a major role in evaluating the judgment too. So, the same video may not be equally suitable for a pair of workers with separate demography. To better understand the other challenges, we took a survey on 45 daily viewers of YouTube of whom 71% were willing to be crowd workers. However, most of them (73%) were ready to provide only a binary response, not the detailed feedback. On an average, they preferred a video segment size of about 140sec and a gap of 37min between the arrival of two successive segments.

Conclusion

The proposed model is a good fit for the YouTube videos because they can be checked as and when uploaded by the users to determine whether they satisfy the safety policies. This model can also be implemented directly in case of streaming videos received from CCTVs for detecting abnormal activities in surveillant area. The approach is also applicable towards detecting the suitability of images and audio, when they get uploaded to other online social media satisfying their policies. Unlike the existing approaches, where videos are blocked with a delay after the users complain, the proposed one urges a real-time action. However, there remain some ethical challenges like managing the risks involved in having contact with potentially harmful media and their impact on the relationship with the workers.

Acknowledgment

The work of Malay Bhattacharyya is supported by the Visvesvaraya Young Faculty Research Fellowship 2015-16 of MeitY, Government of India. All the authors would like to thank the crowd contributors involved in this work.

Molestias odit similique necessitatibus facilis delectus possimus doloremque omnis qui, molestiae rerum repudiandae ipsa quia?