SocialZap: Catch-up on Interesting Television Fragments Discovered from Social Media

Svetlana Kordumova¹, Christoph Kofler², Dennis C. Koelma¹, Bouke Huurnink³, Bauke Freiburg⁴, Joris Kleinveld⁵, Manuel van Rijn⁵, Marco van Deursen⁵, Martha Larson² and Cees G.M. Snoek¹

¹Intelligent Systems Lab Amsterdam, University of Amsterdam, The Netherlands
²Multimedia Information Retrieval Lab, Delft University of Technology, The Netherlands
³Netherlands Institute for Sound and Vision, Amsterdam, The Netherlands
⁴Video Dock B.V., Amsterdam, The Netherlands
⁵Auxilium B.V., Amsterdam, The Netherlands

ABSTRACT

In this paper we present SocialZap, a multimedia search engine that finds the most interesting fragments, zap points, in a television broadcast based on microblog posts and socially tagged photos. The main novelty of SocialZap is the fully-automatic transfer of the learned viewers interest from textual posts to the visual channel, without the need for any manual effort in the process. Once SocialZap finds the zap points, users can easily browse through a television broadcast and directly watch the interesting fragments. Thus, SocialZap adds social experience to watching television.

1. INTRODUCTION

Existing web services [1] allow viewers to watch missed television broadcast later on the web. However, a system that directly suggests the most interesting fragments to watch, based on social media, is, to the best of our knowledge, non-existing. In our SocialZap demo, we analyze data from social media to suggest interesting concepts, zap points, in television broadcast. SocialZap uses textual information from Twitter posts related to a television broadcast of interest, which provide a rich source of information of what viewers find interesting, see Figure 1. The challenge we face is the temporal mismatch between the moment that the user tweets about a concept and the moment it appears in the television broadcast. The tweet-time can radically differ from the appearance-time as viewers either anticipate appearances or continue to tweet about topics that have previously appeared. To transfer the found concepts of interest from the textual to the visual modality, we use state of the art approach [8] which learns concept detectors from social media. We collect Flickr images tagged with the concept names, and we select the most reliable ones as training data for learning. Thus, the proposed system relies totally on social media data, without any manual annotations in the process. The SocialZap pipeline is illustrated in Figure 2, we describe it in more detail in the following section.

2. SYSTEM OVERVIEW

In order to automatically provide zap points to users, SocialZap obtains interesting (visual) topics about the television broadcast and then learns visual concept models to recognize those concepts in an unseen video. SocialZap exploits social information from (1) Twitter posts related to a television broadcast of interest to find the most common terms as concepts of interest and (2) tagged Flickr images to learn visual concept detectors. SocialZap is composed of different modules, which communicate with each other through a common API in a central server, see Figure 3. We describe the separate modules in the following sections.

2.1 Textual Concepts and Social Media

Assuming that viewers write posts about what they find interesting, we exploit social information about the television broadcast gathered from Twitter with the Tweet Analyzer module of SocialZap. The posts are collected by issuing queries with automatically constructed hashtags from the broadcast show name, as #dwdd shown in Figure 2. The tweets that are posted during the time of the broadcast are considered to be associated with the broadcast. Additionally, we exploit closed captions associated with the television broadcast with the Caption Analyzer module. Closed captions are beneficial to collect visually depicted entities if
Figure 2: SocialZap pipeline. Tweets for a particular television broadcast are gathered and visual detectors for the most common terms in the tweets are then learned to find the visual fragments of those concepts as zap points.

there are not enough found in the tweets, and also to denoise tweet-terms by assuring that they also occur at least once in the captions. After obtaining tweets and closed captions, we translate them to English and perform stopword removal. This preprocessing step is necessary to build the basis for later retrieving more accurate training data for selected visual concepts. Each of the terms contained in tweets and closed captions gets assigned a reranking score in the interval \([0, 1]\), which represents both the importance (i.e., popularity) and (visual) depictedness of these terms with respect to the harvested tweets and the (audiovisual) content of the corresponding broadcast. The importance of a term is obtained by building language models [11, 2] for each television broadcast. Terms that are specific to a television broadcast and at the same time mentioned relatively frequently in Twitter posts receive a higher relevance score. In order to quantify the depictedness of a term, we apply

the generic approach from [4]. The overall score of a term is modelled by combining importance- and depictedness-scores in a linear fashion.

2.2 Visual Concepts and Social Media

To transfer viewers interest from the textual modality to visual concepts, we need training data to learn visual concept detectors. In order to detect any possible video concept, a promising line of research is to automatically collect training examples from the web, where many socially tagged videos and images exist [5, 9]. In a recent study, we have shown that tagged images are a better training source than tagged videos for learning video concept detectors [3]. Therefore we use Flickr as source to collect tagged images. Due to variation between users regarding the use of social tags, simply treating all tagged images as positive examples may be problematic [5]. In order to preserve only good examples, we first calculate tag relevance scores per image. We use the multi-feature variant of the neighbor voting algorithm [5]. The images are then ranked in terms of their estimated tag relevance scores. Then, we preserve a top ranked proportion of the examples [3], as shown in Figure 2. As negative examples, we employ negative bootstrap [6], which finds relevant negative examples in an iterative manner. We follow a state-of-the-art bag of visual codes pipeline for feature extraction [10], and learn concept detectors using one-vs-all SVM [7]. We call this module of the framework Video Miner. The described approach implemented in SocialZap showed best performance on the TRECVID 2013 semantic indexing with no annotation task [8], with best score for 31 out of 38 concepts.

Once we obtain the learned concept detectors, we apply them on the broadcasted video frames. For each frame we get confidence scores on how likely the concept appears in the frame. The frames with highest scores are suggested as
zap points of the television broadcast. This module is the Video Indexer.

2.3 User Interface

We created a simple user interface that allows users to catch-up on the interesting fragments of a broadcast, see Figure 4. The interface provides an option to select a specific television show. Once selected, SocialZap lists the most common terms found in the social media posts with respect to the selected broadcast. When the user chooses one of the terms, the video player directly plays the video segment where the significance score for that concept is highest. If some term occurs more than once in the video, all zap points for that term will be listed, as in the case with the term politics shown in Figure 4. As it can be seen from the figure, the concept detector of elections sets the zap point correctly at a relevant frame where this concept appears. However since concept detection is a challenging task and does not always give precise results, we found that timing of tweet and caption terms can help for finding correct zap points.

3. CONCLUSION

In this paper we present the SocialZap system, which finds interesting fragments in television broadcasts based on social media. With SocialZap the viewer’s interest is depicted from Twitter and transferred into the visual channel as zap points. The system provides an easy way to browse through the zap points of the television broadcast, adding a new social dimension to the television experience.

Acknowledgments. This research is supported by the Dutch national program COMMIT; project SEALINCMedia. We thank the students from TU Delft for creating the user interface of Social Zap.

4. REFERENCES