Do sentiments tell a story: Exploring perceptual sentiments in high impact videos

1, 2
King's College London, UK
blah, bleh}@kcl.ac.uk

1. ABSTRACT

This paper tries to explore the art of story telling in the realm of Online Social Networks (OSNs) and online social media. Our work hypotheises that the presence of the well known screen play graph, which directs the over all sentiment of a movie or a drama through time, is very well present in the micro videos posted on the newly available mediums of Vine, instragram and twitter. The In the due course of the work done for the paper, we crawled a popular social media network called Vine for almost 2 months and collected over 12000 unique vine videos and their meta data. We try to take an approach based on perceptual sentiment in social media and hypothesize existence of story lines in perceptual sentiments. We use deep learning tools to detect sentiment values of videos frames and eventually show to a reasonable extent, that perceptual sentiments do follow popular screenwriting theories. The sentiment transitions across these short but high impact videos, do follow certain trends which could be explained from popular screenplay writing theories. The paper also evaluates correlations of individual perceptual sentiments of videos with popularity metrics of the videos. The paper validates presence of generes based on perceptual sentiments of videos and tries to explain them using some popular screenplay techniques.

2. INTRODUCTION

The Art of story telling could be attributed to be one of the most ancient arts. One might give a considerable chunk of credit for the possibility of humans to trace their footsteps across history, to this very art. It could be in the form of neolithic paintings to the egyptian hieroglyphs or among the elaborate epics of Illiad and oddyssey to the elaborate power plays of the Game of thrones, humans have always strived to record or create elaborate plots and stories. The human need of transferring experiences to others in different forms of creative arts, has ever so created the world as interesting as we see it.

The art of story telling is one such creative art, which has spawned and transformed several industries, including the very important entertainment industry. With the invention of the internet and the online social networks, entertainment industry has gone through several rejuvination cycles. The latest of these cycles was powered by the proliferation of services like Youtube, Netflix, Hulu and others like them. Our work in this paper attempts to explore the validity of screenplay and screen writing theories in online social media. More specifically we try to measure perceptual sentiment across a video shared over an OSN, and look for strong evidence in frame sentiments to support generic screen writing theories. We eploy some of the most promising ideas of Visual and perceptual sentiment measurements [2] and couple them with deep learning frameworks for increased generalized performance. Through this work we were able to detect strong clusters of trends of visually percieved sentiments that convey a strong overall story, independently from the actual audio track content.

WWW 2015, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3469-3/15/05.
http://dx.doi.org/10.1145/2736277.2741671.

3. SENTIMENTS IN MEDIA

Sentiments are fundamental part of our day to day social interactions. A face to face social interaction is generally augmented with facial expression, body language and linguistic sentiment to convey the exact meta information. These properties are very human in nature and are mimicked in the social networks as well. Studies like [9] have explored the world of linguistic sentiment in social networks, by comparing several popular textual sentiment analysis methods used for analysing tweets.

When it comes to social media, the analysis becomes complicated. This is because social media involves higher dimensional messages like Videos, audio and Images. Moreover the media shared has a very human centric content. That means the media will involve a lot of faces, poses and affective means of communications. The studies done in [14] show that there has been 900 times increase in the number of selfies over Instagram in just 2 years. Another recent paper [10] states that everyday more than 90 million selfies are taken using just the Android clients out there and are uploaded on Instagram. We collected Vine social network data, which is a popular social network that uses short 6 seconds videos as a medium. In that dataset we found that one in ever three video in the popular videos category contain human faces for more than 60 percent of the frame length. The very human centric nature of the media shared over these networks, make sentiments and human affects an integral part. These mediums When it comes to perceptual sentiments, there are two broad categories that could be explored. The first category looks at the perceptual sentiment evoked by a social media content. The second category talks about the actual latent perceptual sentiment that comes with the context of the content itself. We will discuss about the research problems about both these categories.

3.1 Evoked perceptual sentiment

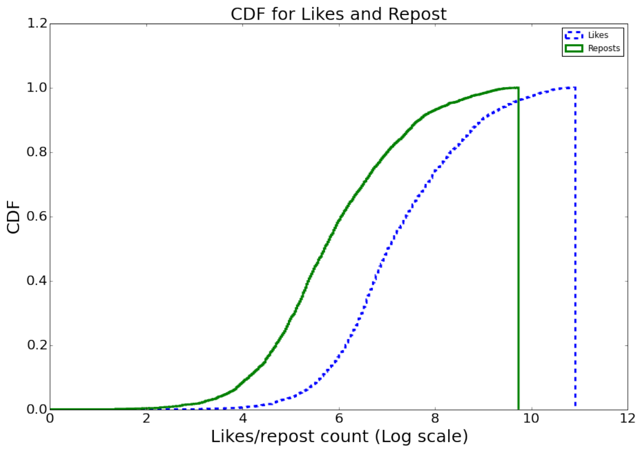

Figure 1: *CDF of Like count and Repost count.*

Several works have done in depth studies using methods like crowdsourcing to understand the different shades of a particular evoked emotion. Works like UrbanGems [1] and StreetScore [13] use crowdsourcing methods to understand degrees of human sentiment evoked because of pictures of real urban neighbourhoods. Sentiments like the feeling of safety and aesthetics are especially hard to quantify and crowdsourcing helps the authors to do some interesting modelling. On the other hand there are papers like [8] by L. Jeni et.al. describe utility of actual facial expression detection for understanding content consumer reaction. Such approaches help us understand the very effect of a particular content on the consumer.

3.2 Latent perceptual sentiment

This approach is what this paper stresses on. By latent perception, we mean the hidden parameters, which are part of the very content. Social networks like reddit have specific sub-reddits that work on appealing to these types media sentiments that evoke emotions like empathy, disgust, contempt and love . One such popular sub-reddit is labelled R/aww which contains images and GIFs that showcase cute animals and animal behaviours. Another one called R/cringe appeals to the sentiments of awkwardness and discomfort by exhibiting videos and Gifs about people in awkward situations. These specific social channels are popular because the content shared over these channels have a certain type of latent sentimental response, which the consumers of these channels resonate with.

Our paper focuses on this part of the story, and tries to survey and benchmark certain state of the art methodologies out there. We also propose certain hybrid approaches, which show that we can attain much better performance if a heuristic approach to combine certain methods is taken.

4. DATASET

We crawled vine for over a month and did a snowball sampling to collect over 12000 videos. The popularity distribution follows as expected a zipf distribution. The videos were then passed through a sampling process, where we sampled one frame from each second of the video lenght. The sampling was chosen purely hurestically. The main aim behind the sampling was to measure the perceptual sentiment as the video progresses. The figure 1 shows the distribution of the crawled vine posts

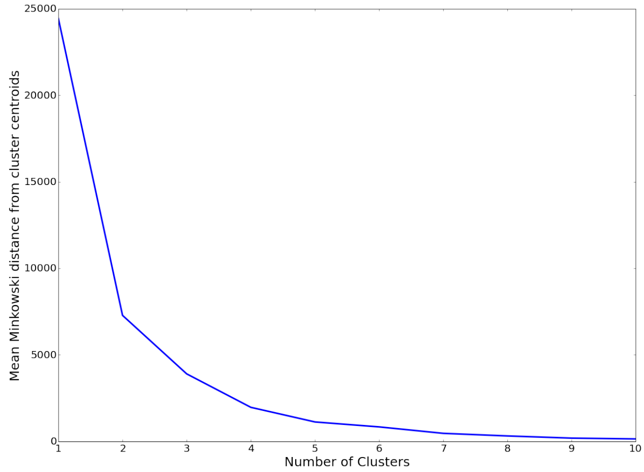

Figure 2: *This graph shows the variation of average minkowski distance of a sentiment vector from a cluster centroid for a given choice of K. The k is varied from 1 to 10.*

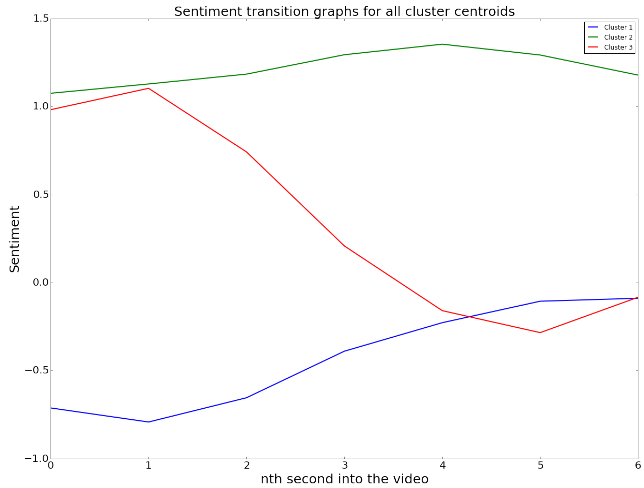

Figure 3: *Sentiment values transitions for the centroids the clusters when K = 3.*

5. SENTIMENTS AND CLUSTERS

We train a deep convolutional network on the Sentibank [**?**] dataset, to classify images in a set of 2079 Adjective noun pairs, each of which have been assigned sentiments using crowd sourced effort. This network gave a top 5 match accuracy for the test dataset of 75%.

The trained visual sentiment detector was then used on chronologically sampled frames from the vine videos collected. We sample 1 image per second for the 7 second long clips and hence now each video was being represented as a vector of 7 sentiment transition values. The main aim of this was to see if users are trying to tell stories in these short vine videos. One of the evidences of this would be clustering of similarly transitioning visual sentiments for videos across the dataset.

6. REFERENCES

[1] ADAM BARWELL, DANIELE QUERCIA, J. C. http://www.cam.ac.uk/research/news/how-to crowdsource-your-happy-space, 2012.

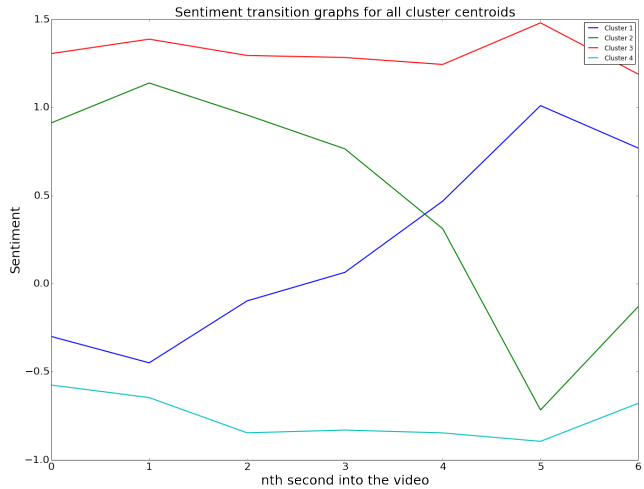

Figure 4: *Sentiment values transitions for the centroids the clusters when K = 4.*

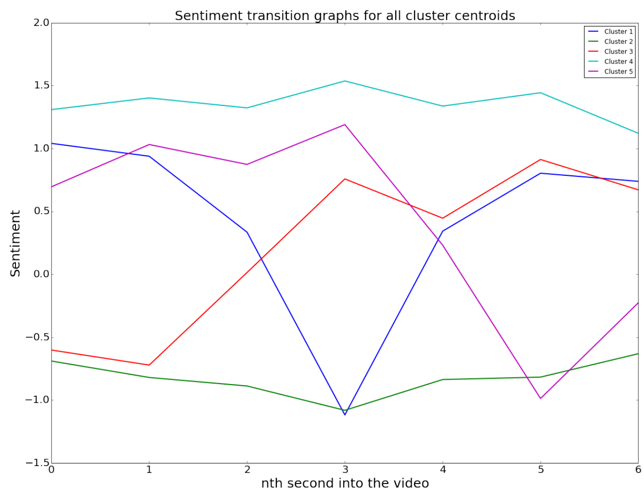

Figure 5: *Sentiment values transitions for the centroids the clusters when K = 5.*

[2] BORTH, D., JI, R., CHEN, T., BREUEL, T., AND CHANG, S.-F. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In *Proceedings of the 21st ACM International Conference on Multimedia* (New York, NY, USA, 2013), MM '13, ACM, pp. 223–232.

[3] BUSSO, C., DENG, Z., YILDIRIM, S., BULUT, M., LEE, C. M., KAZEMZADEH, A., LEE, S., NEUMANN, U., AND NARAYANAN, S. Analysis of emotion recognition using facial expressions, speech and multimodal information. In *Proceedings of the 6th International Conference on Multimodal Interfaces* (New York, NY, USA, 2004), ICMI '04, ACM, pp. 205–211.

[4] DHALL, A., ASTHANA, A., GOECKE, R., AND GEDEON, T. Emotion recognition using phog and lpq features. In *Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on* (March 2011), pp. 878–883.

[5] EKMAN, P., AND FRIESEN, W. V. The facial action coding system. *Consulting Psychologists Press Inc. San Francisco CA* (1978).

[6] HINTON, G., AND SALAKHUTDINOV, R. Reducing the dimensionality of data with neural networks. *Science*, 5786 (July 2006), 504 – 507.

[7] HINTON, G. E., OSINDERO, S., AND TEH, Y.-W. A fast learning algorithm for deep belief nets. *Neural Comput. 18*, 7 (July 2006), 1527–1554.

[8] JENI, L. A., LŐRINCZ, A., NAGY, T., PALOTAI, Z., SEBŐK, J., SZABÓ, Z., AND TAKÁCS, D. 3d shape estimation in video sequences provides high precision evaluation of facial expressions. *Image and Vision Computing 30*, 10 (2012), 785–795.

[9] JOO, J., LI, W., STEEN, F. F., AND ZHU, S. C. Visual persuasion: Inferring communicative intents of images. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2014), pp. 216–223.

[10] KALAYEH, M. M., SEIFU, M., LALANNE, W., AND SHAH, M. How to take a good selfie? In *Proceedings of the 23rd ACM International Conference on Multimedia* (New York, NY, USA, 2015), MM '15, ACM, pp. 923–926.

[11] LIEN, J. J., KANADE, T., COHN, J. F., AND LI, C.-C. Automated facial expression recognition based on facs action units. In *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on* (Apr 1998), pp. 390–395.

[12] LIU, Y., ZHANG, D., LU, G., AND MA, W.-Y. A survey of content-based image retrieval with high-level semantics. *Pattern Recognition 40*, 1 (2007), 262 – 282.

[13] NAIK, N., PHILIPOOM, J., RASKAR, R., AND HIDALGO, C. Streetscore – predicting the perceived safety of one million streetscapes. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on* (June 2014), pp. 793–799.

[14] SOUZA, F., DE LAS CASAS, D., FLORES, V., YOUN, S., CHA, M., QUERCIA, D., AND ALMEIDA, V. Dawn of the selfie era: The whos, wheres, and hows of selfies on Instagram. In *Proceedings of the 2015 ACM on Conference on Online Social Networks - COSN '15* (2015), pp. 221–231.

[15] WANG, W., AND HE, Q. A survey on emotional semantic image retrieval. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on* (Oct 2008), pp. 117–120.