Do sentiments tell a story: Exploring relevence of screenplays and perceptual sentiments in high impact videos

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

1. ABSTRACT

The human need to tell, listen and experience stories in various forms has almost driven various forms of art for thousands of years. From different dramatic dance forms, to puppets, to the modalities enabled by the new age i.e. social media; the basic driver has always been an interesting story. In this paper we explore this very aspect of social media, by trying to propose and validate theories around how videos shared over social media follow popular screenplay patterns and thrive in the digital realm. In the due course of analysis, the paper shows that the most ristricted of videos, also follow story lines and that there are computational ways to classify them. We use human percieved sentiments, and analyse the content in sentiment space. We propose and validate our observations using over 12 thousand videos crawled from popular social media services like Vine and Youtube.

2. INTRODUCTION

The Art of story telling could be attributed to be one of the most ancient arts. 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 has spawned and transformed several industries, including the very important entertainment industry. With progress of technology, entertainment industry has gone through several rejuvination cycles. Starting with plays to television, each technological advancement has created a new form of stories to be presented. The latest of these cycles was powered by the internet with the help of services like Youtube, Netflix, Hulu and Amazon.

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. In the following sections the paper would try to elaborate on Screen play theories, the notion of Sentiments in media, then the paper would introduce you to the dataset we collected, then the techniques of examining the dataset. Finally we would conclude with our findings and discussions. [7]
[1] [2]

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. RELATED WORK

The work done in micro video analysis has been limited. Work by Miriam et.al [11] try to quantify and build on the notion of creativity. Work by [10] use textual sentiments to bring thousands of fiction novels to sentiment space and show that most novels follow 7 salient categories of stories. A paper by Nguyen et.al [9] collected more than 200 thousand micro videos from vine. A work done by Fontanini et.al [3] explore relevence of perceptual sentiments to popularity of a video.

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. The popularity distribution follows a long tail pattern, which is a expected behaviour among a socially influenced dataset.

For training of the Visual sentiment detector, we use the 1 million annotated flicker images opened up to the community by the Sentibank researchers. The accuracy of detector was measured over this training dataset and we baseline our detector's performance based on the mean probability confidence of the detector for the training dataset.

5. SCREENPLAY THEORIES:

Every art, despite being a subjective field, does have certain theories that have emerged over time, which are taught to the novices of the field. These theories propose a boad structure or an approach

[1] www.vanityfair.com/hollywood/2016/02/king-bach-rocketjump-youtube-vine-stars
[2] http://newmediarockstars.com/2015/04/youtubers-viners-attend-the-white-house-correspondents-dinner-gallery/

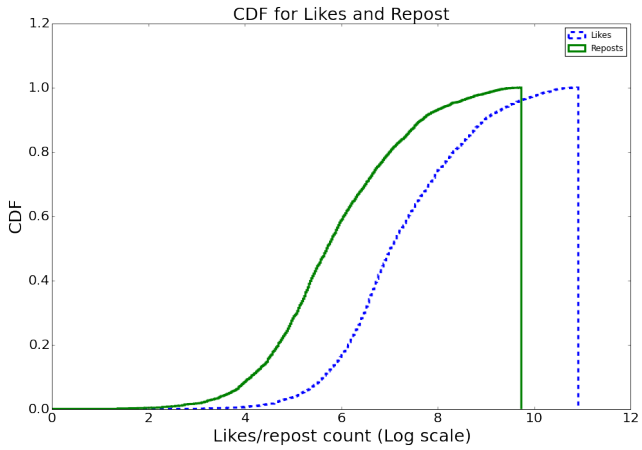

Figure 1: *CDF of Like count and Repost count.*

towards creating the art. Screenwriting is not an exception to this rule. Over decades of excellent cinema and drama, writers and directors have come up with a genral map of how a story is told an how the overall sentiment of a story fluctuates up untill the climax. Let us briefly discuss some of the described categories.

5.1 Aristotle's Three Act Structure

This structure of screenplay is probably the most well studied and has been in the literature since its first mention by Aristotle himself. The structure of the screenplay is generally divided into 3 sections or acts where there is a change in sentiment of the whole story at during every act. This happens generally towards the end of first act because of a situation that poses a challenge or risk to the protagonist of the story. The tension and the protagonist's struggles build up during the second act. Finally the climax of the screenplay which begins towards the end of second act or the beginning of the third, creates a closure for the protagonist by either release of tension or through resolution of a challenge which may come in form of a villian. Throughout the acts for this sort of a structure, the sentiment of the story transitions drastically, going from positive to negative to positive again. Such a transition keeps the viewer engaged and engrossed.

5.2 Hero's journey

Such a screenplay structure is ominipresent across classic stories and typical rags-to-riches stories. The protagonist of the screenplay generally encounters a call to adventure or a proposition of adversity, which he generally accepts and goes on a fluctuationary trajectory of sentiments. The protagonist generally ends up a hero by the climax and the story generally ends on a high sentiment.

5.3 Syd Field's paradigm

Syd Field in his book Screenplay introduced his new concept of paradigms. This is the first theory that explains screenplay writing as a series of plot points and is the most used structure for screenplay out in the modern world. A plot point in a screenplay is a situation in the story that changes the narrative of the story and inevitably the sentiment of the story in some way.

The main contribution of this work, is to suggest that these popular screenplay theories are present in the new age of entertainment videos, which are the micro-videos , found on services like Vine, Instagram and Twitter. Doing so opens up a new paradigm of understanding the field of computational screenplay classification.

6. 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 [5] 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 [12] show that there has been 900 times increase in the number of selfies over Instagram in just 2 years. Another recent paper [6] 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.

6.1 Evoked perceptual sentiment

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 [8] 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 [4] 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.

6.2 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 the second part of the story, and tries to understand relation of this visual sentiment with the ideas of story writing in micro videos.

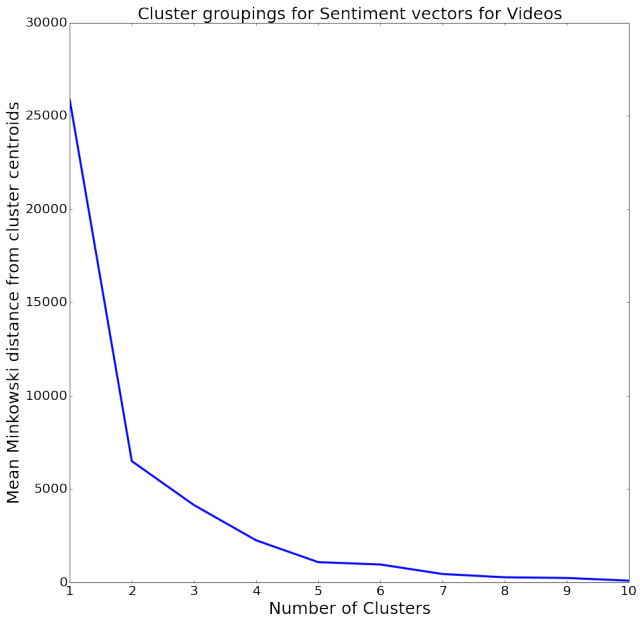

Figure 2: *This graph shows the variation of average minkowski distance of a sentiment vector from a cluster centroid for a given choice of K. The k is varied from 1 to 10.*

7. SENTIMENTS AND CLUSTERS

We train a deep convolutional network on the Sentibank [2] 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.

7.1 Choice of clusters

The clusters were found using the basic k-means algorithm. For this to work we had to represent the videos in a vector representation. For this the videos were first passed through the deep convolutional network and the probability vector which was generated. These vectors were exhamined to We used the elbow point method to find the right choice of clusters to find in our dataset. The metric we use to measure the tightness of clusters was average Minkowski's distance between cluster centroid and the data point vector. The Figure 2 shows the trend of average Minkowski from a cluster centroid as we increase the choice of cluster (k) from 1 to 10. The correct choice ofr k corresponds to the elbow points, or the points where the rate of decrease of average distance changes discontinually. We have 3 such prospective candidates at k = 3 , 4 and 5. The grouping was found to be the tightest at k = 3 and k = 5. Lets discuss what these choices actually mean in terms of sentiment transition values and their relevence to screenplay theories.

7.2 Clusters found

Let us look at the cluster choices individually. Figure 3, 4 and 5 show these trends

7.2.1 k=3

For the choice of k=3, we see the most fundamental grouping of videos by visual perception. There are three distinct clusters with three kinds of perceptual sentiments. The first cluster is made up of all positive sentiments. Such videos mainly find adjective noun pairs which have adjectives like 'Cute' , 'Beautiful', 'Hot' , 'Handsome' et.al. These videos are mostly videos with protagonist being in the frame most of the time and is talking into the camera. This cluster is by sheer count has the biggest share of the pie. The second cluster is the cluster containing mostly negative perceptive sentiments. The adjectives that the network describes the frames by fall in the category of 'Sad' , 'Ugly' , 'messy' et.al. The Third cluster has more of an oscillatory behaviour, where the first half of the video contains positive sentiments and the second negative. Figure 3 plots the centroids of the three clusters. Each curve on the plot represents the second by second sentiment value in the micro-video.

The total number of videos belonging to first type are exactly twice to those belonging to the second type. This points towards a fact that users like to share fundamentally feel-good videos.

7.2.2 k=4

For k=4, the grouping is weakest. But the signature of the two prime clusters of all negative and all positive sentiments still exist. The most populous two clusters are the all positive sentiments, followed by the all negative. To explain the other two clusters, we have to move the the choice of k=5, where in the screenplay theories begin to correlate.

7.2.3 k=5

This choice of k had the second best grouping amongst the 3 candidate choices. The two big clusters still follow the theme of mostly positive and mostly negative perceptive sentiments. But we see three interesting trends which follow three possible explaination based on the screenplay theories

We can first try analysing the sentiment trends using the popular Aristotle's three act play structure as discussed in Section **??**. In this structure generally all stories can be fragmented into three segments. In each segment generally the story undergoes a turmoil or a drastic change in sentiment. In case of cluster C3, you can see these changes go towards a positive sentiment trajectory.

8. REFERENCES

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

[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] FONTANINI, G., BERTINI, M., AND DEL BIMBO, A. Web video popularity prediction using sentiment and content visual features. In *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval* (2016), ACM, pp. 289–292.

[4] 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

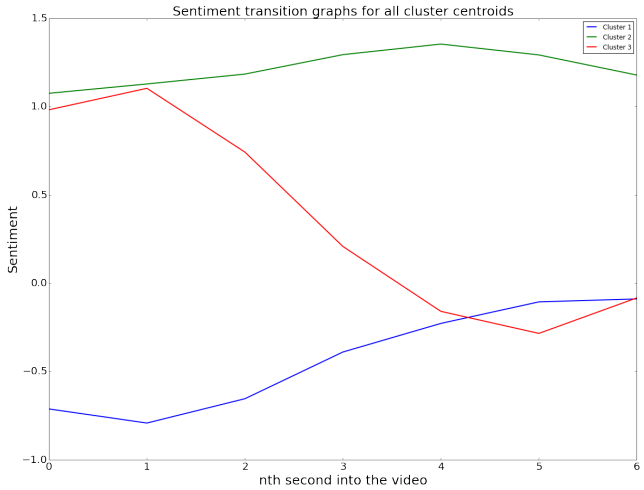

Figure 3: *Sentiment values transitions for the centroids the clusters when K = 3.*

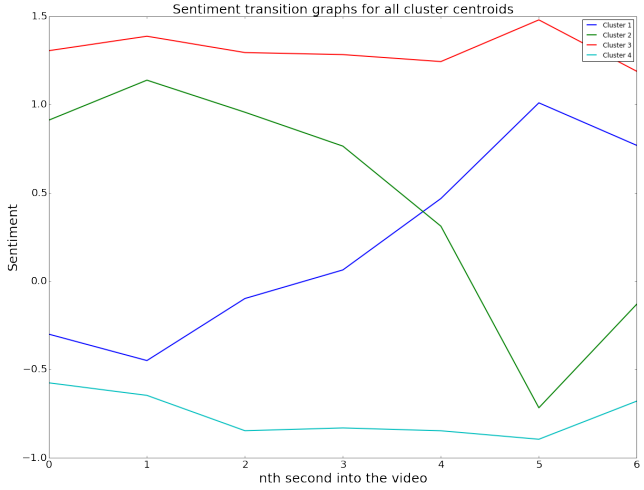

Figure 4: *Sentiment values transitions for the centroids the clusters when K = 4.*

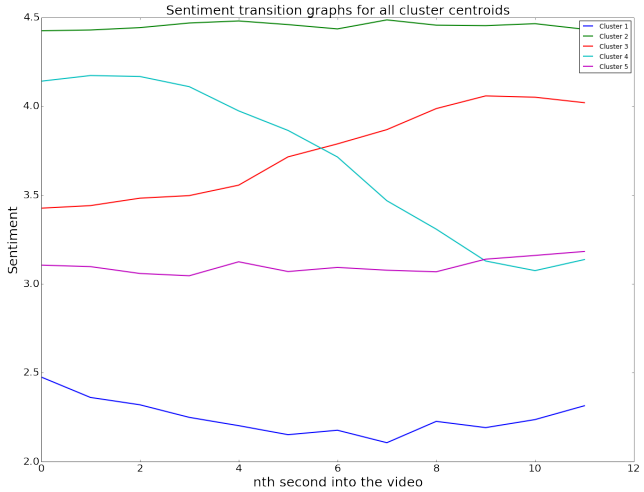

Figure 5: *Sentiment values transitions for the centroids the clusters when K = 5.*

evaluation of facial expressions. *Image and Vision Computing 30*, 10 (2012), 785–795.

[5] 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.

[6] 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.

[7] MARCOS-RAMIRO, A., PIZARRO, D., MARRON-ROMERA, M., AND GATICA-PEREZ, D. Let your body speak: Communicative cue extraction on natural interaction using rgbd data. *IEEE Transactions on Multimedia 17*, 10 (Oct 2015), 1721–1732.

[8] 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.

[9] NGUYEN, P. X., ROGEZ, G., FOWLKES, C., AND RAMAMNAN, D. The open world of micro-videos. *arXiv preprint arXiv:1603.09439* (2016).

[10] REAGAN, A. J., MITCHELL, L., KILEY, D., DANFORTH, C. M., AND DODDS, P. S. The emotional arcs of stories are dominated by six basic shapes. *arXiv preprint arXiv:1606.07772* (2016).

[11] REDI, M., O'HARE, N., SCHIFANELLA, R., TREVISIOL, M., AND JAIMES, A. 6 seconds of sound and vision: Creativity in micro-videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2014), pp. 4272–4279.

[12] 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.