Understanding popularity of vines

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

1. ABSTRACT

Users on social content-sharing websites are faced with a plethora of choice today, and make quick decisions about resharing or liking posts. How should content creators capture viewers' attention in this age of information overload? Does quality of content matter? Or is there greater support for "rich gets richer" theories, which suggest a self-perpetuating phenomenon where content from users with large numbers of followers stands a greater chance of becoming popular? To the extent that quality matters, what aspect of the content is critical to ensuring popularity? We examine these questions using a snapshot of nearly all videos uploaded to vine over a 8 week period. We find that although social factors do affect popularity, content quality becomes critical at the top end of the popularity scale. Furthermore, using the temporal aspects of video, we verify that decisions are made quickly, and that first impressions matter more than deeper impressions.

2. INTRODUCTION

In an age of information overload, exactly how

3. RELATED WORK

The work done in micro video analysis has been limited. Work by Miriam et.al [17] try to quantify and build on the notion of creativity. Work by [16] 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 [14] 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, but the work done was on youtube viral videos, which have a much richer composition and structure. The problem of understanding what makes a visual media stick, has been a difficult one to solve. There are a few approaches to understand the aesthic and memorability aspects of an image [7] [2] [9]

[**?**] –> Likes does not propagate quickly through Flickr social network (so quality is important)

[**?**] –> no correlation between age and popularity; most photos gain most likes in the first week.

[**?**] –> performs quite well, predicting popularity through text annotating features.

4. BACKGROUND AND METHODOLOGY

4.1 Introduction to Vine

Vine[1] is a video sharing platform owned by Twitter, where users can create and share videos called "vines" up to 6 seconds long, a constraint which was designed with the goal of inspiring "creativity"[2]. Vine is primarily used as a mobile app, although a Web interface as well as an Xbox Live interface is available to view the videos. Users may create vines and upload them to the platform (typically through the Vine mobile app), view vines created by others, and follow other users whose vines they find interesting.

Users see vines created by others in one of several tabs: The home tab shows a personalised and social feed of videos created by those that they follow. The explore tab shows 20 feeds and channels. Eighteen of the channels are category-specific, such as 'Comedy', 'Music', 'Animals', 'Weird', 'Sports', 'Arts' etc; and appeal to different kinds of users based on their specific interests. Vines are assigned to specific channels by their creators. The remaining two channels are termed by Vine as 'popular-now' and 'on-the-rise', and are curated channels containing videos that have proven to be of of wider appeal to the entire vine population.

As with other modern content-sharing platforms, there is a social aspect: viewers can follow others whose vines they find interesting, 'like' or share interesting vines by 'revining' (i.e., reposting) the vines, and commenting on them. The numbers of likes, revines and comments are a measure of the popularity of a video. Uniquely, vine plays videos in a loop, going back to the beginning after reaching the end. Loops are repeated as long as the focus is on that video (e.g., video is active on the mobile phone screen if using the vine app, or mouse is on the video in the web version). Thus, letting a video play for more than one loop can be a sign of engagement. Therefore the platform also tracks and reports in real time the aggregate number loops across all users.

4.2 Dataset description

Dataset	Vines (total)	Loops (median)	Reposts (median)	Likes (median)
POP12K	11448	318566	2173	7544
UNPOP120K	122327	80	0	2

Table 1: Table

[1] http://vine.co
[2] http://blog.vine.co/post/55514427556/introducing-vine

The data[3] used in this paper is summarised in Table 1, and were collected in two phases as described below:

4.2.1 Popular videos dataset

First, we collected ≈12,000 videos which have been marked by vine as ‘popular’, by tracking the ‘popular-now’ channel[4] over a three week period in Dec 2015, and downloading all videos and associated metadata once every six hours, and removing any overlapping videos from the previous visit. The crawling period was chosen to ensure that consecutive crawls have an overlap of several videos, and this sufficed for all visits made to the website during the data collection period; thus the dataset we collected is a complete collection of all ‘popular-now’ vines during the 21 days under consideration.

Vine does not disclose the algorithm used to mark a vine as popular; yet we observe (see Table 1) orders of magnitude more loops, reposts and likes in the popular-now dataset than in the channel dataset. Thus we believe that the algorithm used by vine to select vines for the ’popular-now’ channel is strongly affected by the numbers of loops/revines/likes. Note that the numbers of loops etc. were collected at the time of crawl, within a maximum of six hours of being posted on the ’popular-now’ channel, which limits the possibility that the counts increased *as a result* of being featured on the popular-now channel. In the rest of the paper, we use the counts in the popular-now dataset to calibrate the definition of ‘popular’. While there is a possibility that this is a biased proxy for global popularity, it nevertheless provides a baseline against which to compare all videos.

4.2.2 All channel videos dataset

In the second phase, we collected videos accessible from each of the 18 global vine channels or categoriesover a period of **8 weeks** from **Aug XX to Oct YY 2016**. Again, a crawling period of six hours was chosen for consecutive visits to the same channel, and the 100 most recent vines were fetched with each visit. As shown in Fig. 1, the vines returned has a significant overlap with vines fetched from the previous visit. Thus we believe that our dataset captures nearly all videos uploaded to vine and assigned to a channel. The only exception is the extremely popular comedy channel, for which we nearly always find more than 100 new videos (we only download the 100 most recent videos for the comedy channel). In total, this results in a dataset of ≈120,000 videos. We track the loop, revine and like counts over time, periodically updating each video’s counts every three days until the end of the data collection effort.

Note that while we obtain nearly all videos across the channels, our dataset does *not* capture *all* videos uploaded to vine – vine creators do not need to assign a video to a channel, and we do not discover any vines not in channels. We use channels to restrict ourselves to vines which get exposed to a reasonably global audience of those interested in a topic category, and therefore to vines that have a higher potential for garnering high like/revine/loop counts.

The popularity distribution of the whole dataset follows as expected a zipf distribution. The Fig. 1 shows the distribution of likes and repost counts of the collected videos on al og scale. Videos with 0 likes or reposts were given a marginal 1 like to avoid undefined logarithms.

Over the course of this study, we also used several third party datasets, to corroborate different properties of vine videos, with datasets previously studied, ranked and evaluated by other research groups. For comparing the aesthetic qualitiy of our videos, we used

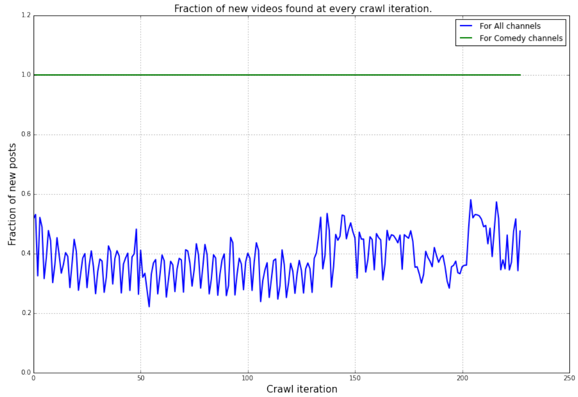

Figure 1: Fraction of new videos since last visit amongst the videos whose metadata was fetched by requesting for the 100 most recent videos from each channel is consistently less than 1, suggesting that the dataset contains nearly all videos from most channels. The only exception is the comedy channel, which consistently has more than 100 new videos (thus fraction of new videos is nearly always 1).

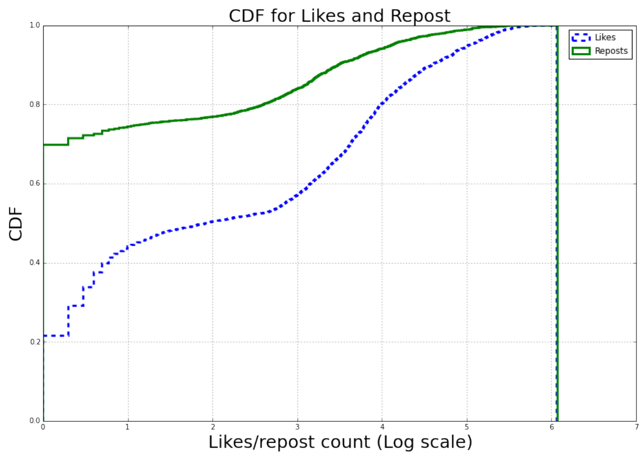

Figure 2: *CDF of Like count and Repost count.* **The values are normalized and on a Logarithmic scale. As expected from a long tail distribution of metrics like popularity, the dataset has a lot of videos with zero likes and reposts. Such videos are synthetically given one like and repost to avoid undefined values**

the dataset provided by Datta et.al [2] which provided rated images , rated by a crowdsourced activity, on the scale of 0 to 7. Each image has more than 15 ratings, and the median ratings of these images were considered for our work. For comparison with static images, we use the popular MIR-Flickr dataset [6] which is an open source dataset of over 25,000 images crawled from the Flickr image service. For our work with frame sentiments, we use the dataset provided by the work of sentibank [8] [1] for baselining the performance of our detector for detecting frame sentiments in vine. Finally for doing the exploratory work with facial expressions, we used the dataset opened up to the public for the facial expression recognition competition 2013 [4] to train and test our convolutional network

5. FEATURES

[3]All datasets will be made available for non-commercial research.
[4]https://vine.co/popular-now

Over all the challege was to understand what makes a vine popular. And for that the there was a need to explore correlations of all the possible abstract midlevel features made available to us because of the rapid development in the fields of Deep machine learning. The main important contribution of machine learning is the ability of computationally extracting abstract higher level representations, which vaguely represent human perception.

5.1 Low level Aesthetic Features

: There are some well known computationally evaluatable aesthetic features which have been recognized as heuristics for good photography. Examples include Rule of thirds, Sharp Pixel proportion, Contrast, Simplicity, Left- Right symmetry etc [19]. The parameters basically compute perceptual features of an image based on well know heurestic rules set by photographers. Some of the detailed references of the features we use are

5.1.1 Contrast

Contrast is basically dissmilarity between pixel(colour) values in a picture. It is a good aesthetic measure to understand how the photographer or creator of a visualc content has used the range of colour values to his advantage. This measure does not always reflect the aesthetic quality of an image, but with other features like sharp pixel proportion, can be a good approximation. For the sake of our study, we use Weber contrast, which is defined as

$$F_{weber} = \sum_{x=width} \sum_{y=height} \frac{I(x,y) - I_{average}}{I_{average}} \quad (1)$$

5.1.2 Simplicity

Simplicty of composition of a photograph is a distinguishible factor that directly correlates with professionalism of the creator [10]. We use simplicity definition as defined in [19] to calculate the ROI segment simplicity and Luo simplicity [13]

5.1.3 Rule of Thirds

This feature deals with compositional aspects of a photograph. Several papers including [19] study this feature and hence we use this as one of our aesthetic features. This feature basically calculates if the object of interest is placed in one of the imaginary intersection of lines drawn at approximate onethird of the horizontal and vertical postions. This is a well known aesthetic guideline for photographers.

5.1.4 Sharp Pixel Proportion

Out of focus or blurry photographs are generally not considered aesthetically pleasing. In this feature we measure the proportion of sharp pixels compared to total pixels. To do so we have to transform the image from intensity domain to frequency domain, and then count the total number of pixels which surpass the shapness criterion. We choose the criterion of sharpness in frequency domain to be 2 from [19]. The processing of the images was done using a tool called OpenIMAJ [5]

5.1.5 L-R Balance

Difference in intensity of pixels between two sections of an image is also a good measure of aesthetic quality. In non-ideal lighting conditions, images and videos tend to be over exposed in one part and correctly exposed in other. This is generally a sign of amature creator. To capture this we compare the distribution of intensities of pixels in the left and right side of the image. The distance between the two distributions is measured using Chi-squared distance.

Table 2: List of Aesthetic parameters computed for highly rated aesthetic images, Popular videos and unpopular videos. Most parameters have no bias towards either popular or unpopular videos

Parameter	Aesthetic Images		Popular Vines		Unpopular Vines	
_	Mean	Median	Mean	Median	Mean	Median
Color Contrast	51.05	30.22	29.88	16.43	20.23	8.83
Intensity Balance	0.11	0.08	0.16	0.13	0.17	0.14
Luo Simplicity	0.009	0.005	0.013	0.012	0.015	0.014
Sharp pixel proportion	0.103	0.098	0.090	0.085	0.089	0.081
Image Saturation	0.943	0.974	0.672	0.678	0.615	0.646
Avg. Brightness	0.148	0.141	0.137	0.130	0.139	0.124
Rule of Thirds	0.879	0.899	0.883	0.883	0.878	0.882
ROI Proportion	0.316	0.089	0.175	0.112	0.165	0.110

5.1.6 Naturalness

This is a very heuristic property if an image that tries to gauge the degree of correspondence of images to the human perception. We first convert the image from RGB to HSL colour space which is proved to be closer to human perception of colours. We then group pixels using a heuristic rule that chooses pixels corresponding to natural objects like skin, grass, sky, water etc. This is done by choosing pixels which have L $\in$ [20 , 80] and S > 0.1. The final naturalness score is calulated by finding the weighted average of all the groups of pixels. [20]

5.1.7 Colourfulness

This is measure of an image's difference against a pure Gray image. It calculated as specified in [19]

For baseline and comparison, we compare the features with images taken from the dataset from photo.net [2]. We only choose images with median ratings of 6 or above on aesthetic scale of 0 to 7. Because of the very nature of Vine videos, it was possible to sample 6 images , one for each second of the video, to get a good approximation of the aesthetic quality of the whole video. So in our processing pipeline, we sample one image per second from the 6 second long vine, and then take a median score of the aesthetic parameter across the sampled image. This score is assigned to the whole video to signify the value of that particular aesthetic parameter.

5.2 Presence of Faces

One important aspect of micro videos is the presence of user as an actor in the video. When you look at viral vine videos, most videos seem to have a lead actor performing a skit. The hypothesis here was that vine has become a social media network, where actors have gained prominence and become a reason for popularity. So we did a small experiment, where we sample one image every second from all the videos collected. Then we calculate the percentage of frames across each video which contained at least one face in it. We use the well tested Viola Jones dectector for frontal and profile face detection. [18]. When we plot the CDF of these percentages for popular against unpopular videos we see considerably higher population of popular videos to have high number of face image percentage 3.

5.3 Frame Sentiments

The crawled vine dataset was sampled and processed using the Multi Lingual sentibank detectors [8] which expresses visual sentiment of video feames on the scale of 1 to 5, 1 being negative and 5 being positive sentiments. To make use of this framework we randomly select about 12000 videos from the collection of 80k plus videos having both popular and unpopular vines. Then we sample these videos for frames twice every second. This creates a time se-

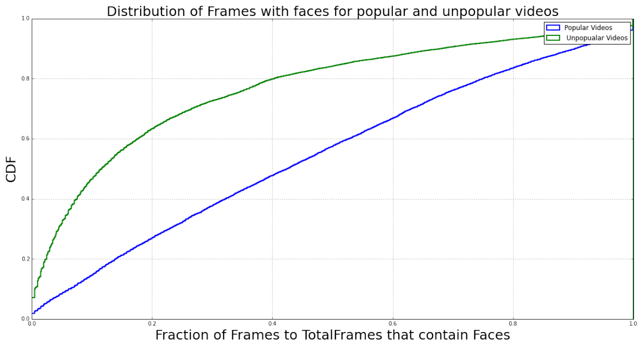

Figure 3: *CDF for popular and unpopular videos. The CDF signifies the cumulative distribution of percentages of face containing frames in a vine video. The observation here is popular videos tend to have higher percentages than unpopular videos*

ries of video frames which could be now fed into the deep neural network for estimation of sentiments. These sentiments are basically vectors of length 12, (vine videos can be at most 6 seconds long) and there are 12000 such vectors. We can now do some statistical analysis on this N x 12 matrix

5.4 Audio Features

Along with aesthetic and perceptual sentiment features, audio is a big part of the Vine clip. We decided to use features that resonate more with the perceptual side of the analysis than the low level. We use 6 features that signify [17] perceptual attributes like loudness, rhythemical features, roughness etc. We extract these perceptual features using open source tools and incorporate them in the list of features we use for analysis [11] [12].

5.5 Social Features

To incorporate the effects of the social visibility for a post, the follower count and the past post count of the user who posts a particular video is used along with all the percieved features.

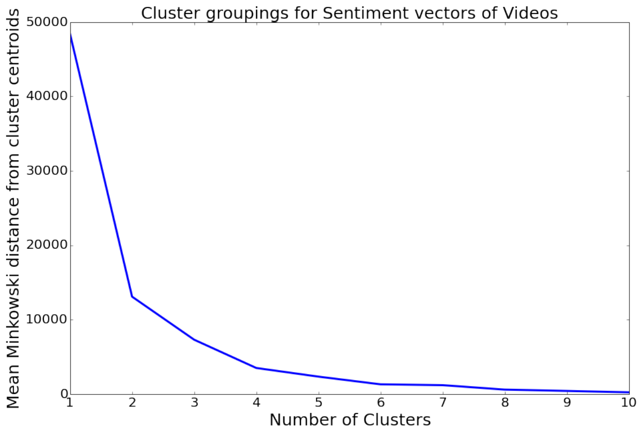

Figure 4: *After doing a simple dimentional reduction of the frame sentiment vectors of all the videos using Principal component analysis, they were analysed for existence of clusters, using the elbow point methods for mean Minkowski distance. We found that the best grouping exists at K=4*

6. INSIGHTS FROM THE FEATURES

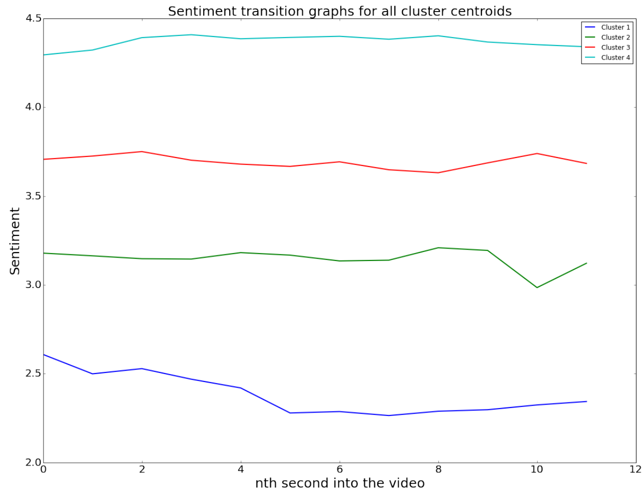

Figure 5: *Plot of frame sentiment vectors, of the 4 centroids of the clusters found. The vine videos tend to have a constant sentiment structure at 4 distinct sentiment levels*

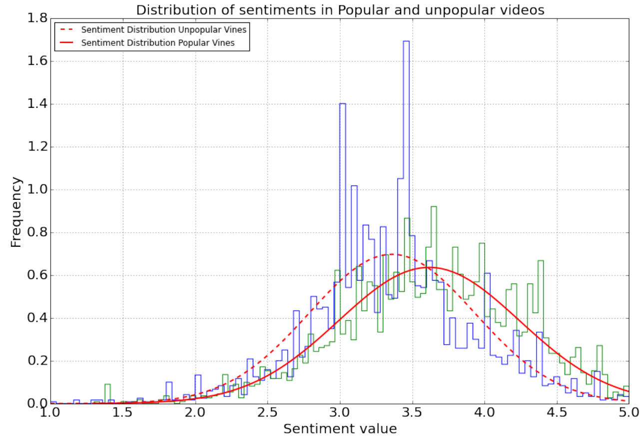

Figure 6: *Distribution of sentiment values for Popular and unpopular videos. The distributions follow a Gaussian like curve, but Popular videos tend to have more positive sentiments than Unpopular*

From the above explored visual, higher level features, it seems that there are a few aesthetic and affective properties of a video, that seem to be over expressed in a popular vine. The idea of vine is a micro video website, but the interpretation and projection of these microvideos as an art from is kept completely open to the users. It is in such sandboxes, interesting art is created. YouTube began with a similar motive and it has now evolved into a gargantuan platform for showcasing different artforms that trancent national bountries, languanges generes. Vine takes this one step further and ristricts the time. This can be a limiting factor or can be used in a peculiar way. Clearly it is the former as some of these videos get likes, shares and loops on par with famous viral videos. The question is what are these unique styles of the users that have allowed them to breach this ristriction.

From the initial work done on aesthetic and affective aspects of a vine video in our paper, it seems the popularity is less of a function of aesthetics, but more of the content. Apart from colour contrast,

Feature Name	Dimensions	Description
Mean Sentiment	1	Mean of sentiments detected using sentibank [15]
Contrast	3	Frame contrast calculated using Webber, color and RMS techniques
Simplicity	2	Image simplicity calculated by two methods [19]
Naturalness	1	A measure of "Naturalness" of a frame
Colourfulness	1	A measure of colourfullness that describes the deviation from a pure gray image
Hue Stats	2	Hue mean and variance which signifies the range of pure colours present in the image
LR balance	1	The Chi squared distance between the histogram of Left and Right side of image pixels.
Object Saliency	2	Measure of prominance given to salient objects. Includes Rule of thirds and ROI proportion
Image brightness	3	Features signify brightness of the image. Includes average brightness, saturation and saturation variance
Image sharpness	2	Features signify how sharp an image is. Includes sharpness variance and sharp pixel proportion
Audio Rhytmical Features	2	Onset rate and zero crossing rate which talks about rhythmic component of track [11]
Loudness	2	Overall energy and average short time energy which signifies loudness of the track [11]
Mode	1	Musical mode of the audio tract (major or minor). [11]
Roughness	2	measure of dissonance values between all peak pairs in the track [11]
Face Percentage	1	Percetage of frames in a video, which have been tested positive for atleast one face [18]
Social Features	2	Number of followers and past number of posts uploaded

Table 3: Dimensionality and description of features used for training the Classifier

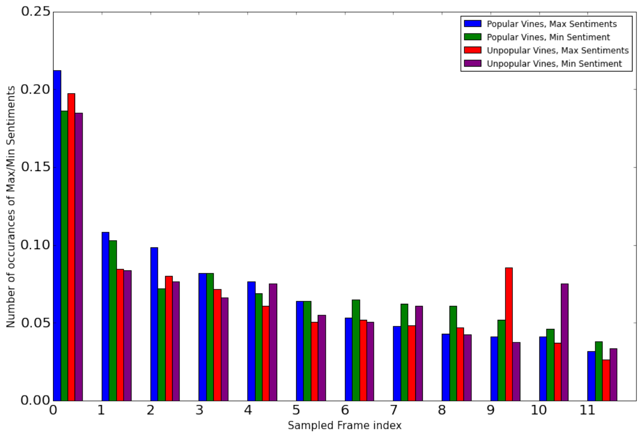

Figure 7: *A graph showing frequency of occurance miximum and minimum frame sentiment in a video against which half second period the frame was sampled from. Interestingly for both popular and unpopular videos, the maximum and minumum sentiments most frequently occur in the first second of the video*

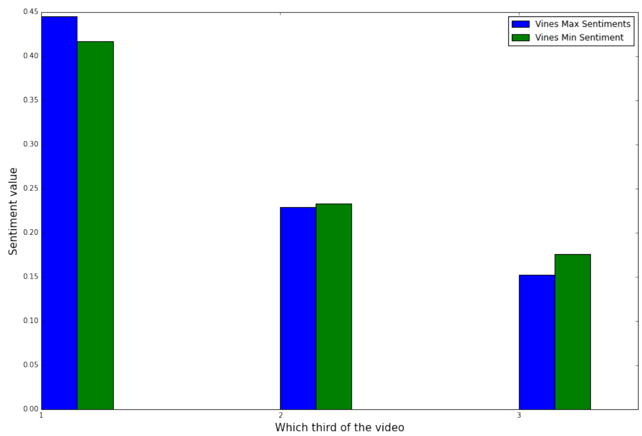

Figure 8: *Frequency of occuranceo of maximum or minimum sentiment. For this graph, the video is considered in thirds, and the frequency of occurance of both max and min frame sentiments is plotted.*

none of the 9 major aesthetic parameters evaluated for both popular and unpopular vines, showed any form of over expression 3. Most parameters either were at par aestheic baseline images or fell way behind. But between the unpopular and popular set, there was hardly any formidable statistical difference.

One major property that shows its presence was the presence of a human face 3. Popular videos show a very high bias towards videos with human faces in them. This says a something about why the micro video service of vine has been transformed more towards a micro vlogging service. Users tend to make micro-skits with a story or a shock value, which delivers humour in a very unconventional way.

Frame sentiments also seem to have some impact on the popularity of the video. Popular videos tend to have a shift towards positive sentiments compared to unpopular videos as seen from 6. Moreover Vine videos in general, tend to set up the mood of the video in the first onethirds of the video. You will see observe either the maximum or the minimum frame sentiment of the video within the first onethirds of the video.

7. DESIGNING A CLASSIFIER

The study aimed to understand the relevance of different kinds of features for the popularity of the video, and gain some empirical evidence of the effects involved in the process of popularization of a micro video. Using the exhaustive data we have crawled, we train a Random Forrest classifier on the 28 perceptual and social features. We sample 10% of the post from the UNPOP120K dataset. As the popularity is a long tailed phenomenon, we also sample an additional 5000 posts from the POP 12K gold standard dataset. This allowed us to assemble a balanced dataset of nearly 40% popular and 60% unpopular videos. These videos were then sampled and processed to extract the 28 dimentional feature vector, that described the video in terms of both perceptual, aesthetic and social features.

After sampling from the two datasets we have features from 17000 videos and their corresponding like, loop and repost counts. Further a random forest classifier is trained on 70% of the resulting feature set and tested on the remaining 30% for performance. Once the classifier is trained you can commend about the impact of each of the feature on the overall classification process by looking at the classifier coefficients for each feature. Hence the training

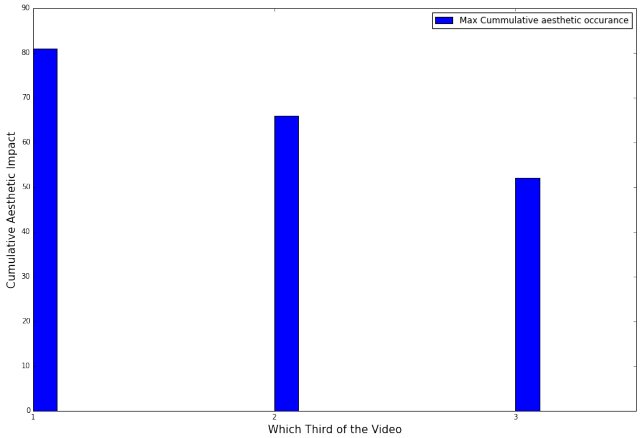

Figure 9: *Plot of cumulative aesthetic impact of each third of the video. For this plot the videos were sampled at one frame a second, and aesthetic features were calculated at every second. Finally the features were*

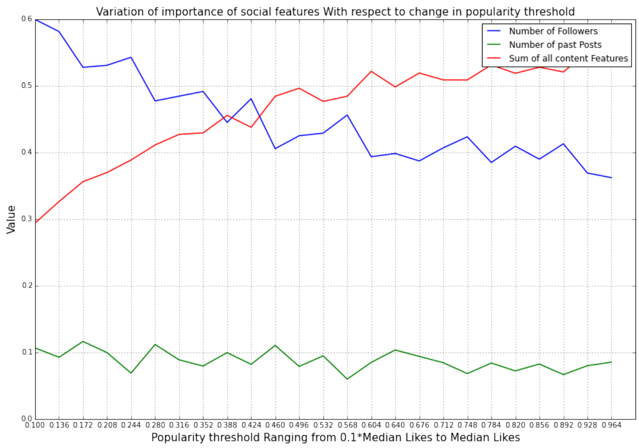

Figure 10: *A plot of contribution of social features against all the perceptual features combined. The influence is calculated by training the classifier and looking at the coeeficients of the final function. The values are generated by interating the training and testing process for different thresholds for the definition of a popular post. The x axis signifies the threshold value of Likes at which a video is labelled to be popular. It starts from the median of the UNPOP120K and ends at median of POP12K.*

process was repeated several times for the dataset, but with different thresholds for popularity. The threshold ranges from median of the whole UNPOP120K to the median of POP12K dataset. This helps us understand a diverse range of popularity definitions which are as liberal as labelling one in every 3 videos as popular to as conservative as labelling only 10% of the videos as popular. Fig 10 shows the transition of impact of social features as compared to the aesthetic and perceptual features. It is interesting to note that the more picky you become in regards to definition of popularity, the higher importance is placed on aesthetic and perceptual features. This says something about the nature of the micro video service. It is easier to get into the club of popular videos if you have sizable followers, but to reach the top of the popularity, your content needs to be good

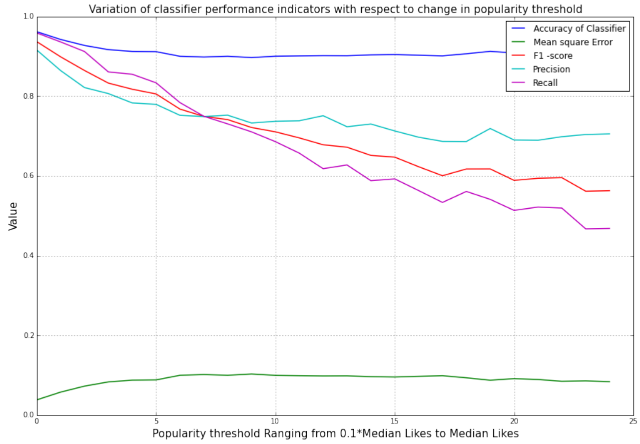

Figure 11: *The plot shows varying values of Precision, Recall, Accuracy and F1-score across the classifier traning iterations*

8. REFERENCES

[1] 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.

[2] DATTA, R., LI, J., AND WANG, J. Z. Algorithmic inferencing of aesthetics and emotion in natural images: An exposition. In *2008 15th IEEE International Conference on Image Processing* (2008), IEEE, pp. 105–108.

[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] GOODFELLOW, I. J., ERHAN, D., CARRIER, P. L., COURVILLE, A., MIRZA, M., HAMNER, B., CUKIERSKI, W., TANG, Y., THALER, D., LEE, D.-H., ET AL. Challenges in representation learning: A report on three machine learning contests. In *International Conference on Neural Information Processing* (2013), Springer, pp. 117–124.

[5] HARE, J. S., SAMANGOOEI, S., AND DUPPLAW, D. P. Openimaj and imageterrier: Java libraries and tools for scalable multimedia analysis and indexing of images. In *Proceedings of the 19th ACM international conference on Multimedia* (New York, NY, USA, 2011), MM '11, ACM, pp. 691–694.

[6] HUISKES, M. J., AND LEW, M. S. The mir flickr retrieval evaluation. In *MIR '08: Proceedings of the 2008 ACM International Conference on Multimedia Information Retrieval* (New York, NY, USA, 2008), ACM.

[7] ISOLA, P., XIAO, J., TORRALBA, A., AND OLIVA, A. What makes an image memorable? In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2011), pp. 145–152.

[8] JOU, B., CHEN, T., PAPPAS, N., REDI, M., TOPKARA, M., AND CHANG, S.-F. Visual affect around the world: A large-scale multilingual visual sentiment ontology. In *Proceedings of the 23rd ACM international conference on Multimedia* (2015), ACM, pp. 159–168.

[9] 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.

[10] KE, Y., TANG, X., AND JING, F. The design of high-level features for photo quality assessment. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)* (2006), vol. 1, IEEE, pp. 419–426.

[11] LARTILLOT, O., AND TOIVIAINEN, P. A matlab toolbox for musical feature extraction from audio. In *International Conference on Digital Audio Effects* (2007), pp. 237–244.

[12] LAURIER, C., LARTILLOT, O., EEROLA, T., AND TOIVIAINEN, P. Exploring relationships between audio features and emotion in music.

[13] LUO, Y., AND TANG, X. Photo and video quality evaluation: Focusing on the subject. In *European Conference on Computer Vision* (2008), Springer, pp. 386–399.

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

[15] PAPPAS, N., REDI, M., TOPKARA, M., JOU, B., LIU, H., CHEN, T., AND CHANG, S.-F. Multilingual visual sentiment concept matching. *arXiv preprint arXiv:1606.02276* (2016).

[16] 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).

[17] 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.

[18] VIOLA, P., AND JONES, M. J. Robust real-time face detection. *International journal of computer vision 57*, 2 (2004), 137–154.

[19] YEH, C.-H., HO, Y.-C., BARSKY, B. A., AND OUHYOUNG, M. Personalized photograph ranking and selection system. In *Proceedings of the 18th ACM international conference on Multimedia* (2010), ACM, pp. 211–220.

[20] ZHONG, C., KARAMSHUK, D., AND SASTRY, N. Predicting pinterest: Automating a distributed human computation. In *Proceedings of the 24th International Conference on World Wide Web* (New York, NY, USA, 2015), WWW '15, ACM, pp. 1417–1426.