Predicting Popularity of Images Over 30 Days

Amartya Dutta
Undergraduate Student
Department of Computer Science and Engineering
Indian Institute of Information Technology Guwahati

Dr. Ferdous Ahmed Barbhuiya
Associate Professor
Department of Computer Science and Engineering
Indian Institute of Information Technology Guwahati

Abstract

The current work deals with the problem of attempting to predict the popularity of images before even being uploaded. This method is specifically focused on Flickr images. Social features of each image as well as that of the user who had uploaded it, have been recorded. The dataset also includes the engagement score of each image which is the ground truth value of the views obtained by each image over a period of 30 days. The work aims to predict the popularity of images on Flickr over a period of 30 days using the social features of the user and the image, as well as the visual features of the images. The method states that the engagement sequence of an image can be said to depend on two independent quantities, namely scale and shape of an image. Once the shape and scale of an image have been predicted, combining them the predicted sequence of an image over 30 days is obtained. The current work follows a previous work done in the same direction, with certain speculations and suggestions of improvement.

1   Introduction

In recent times, there has been a great interest in trying to find out the popularity of an image or a post. Such information holds importance for a lot of purposes. This could vary from social media campaigns, marketing, raising awareness on an issue or recommendation systems. In context of social media campaigns, the social dynamics of a post will inform a user about how long the post continues to engage the viewers. Thus, when that post becomes obsolete, another one can be posted to maximise the audience engagement.
In Khosla et al. [1],the authors stated that using the social features of the user

who posts the image, the social features of the image and the visual features of an image, a popularity score could be assigned to each image. Accordingly, they defined a normalised log function that would define the popularity score of an image as

$$score_i = log(c_i/T_i + 1) \qquad (1)$$

where, c_i is the engagement score (here, the total number of views since upload) and T_i is the time since upload.

2   Related Work

Khosla et al. [1], rightly showed through their equation how the popularity of an image varied over time. A particular image remains usually popular during the initial time of upload but with time, it's popularity drops. However, the equation compared popularity of two images without considering their engagement dynamics. For instance, an image uploaded not very long ago, having a greater number of views would have a higher popularity score than another which was posted much ago but now had almost the same number of views. This was even stated in Valafar et al.[7],that the number of views increases for the first few days and then becomes stagnant slowly with time. Thus Ortis et al. [2,3] stated that the log function as suggested by [1] does not take into consideration the dynamic of the engagement of a post. This way an older photo is directly compared to a newer photo and thus punished by giving it a much lower popularity score.

In [2,3], the authors therefore stated that the engagement dynamics of an image is also important and needs to be considered while predicting the popularity of images. Thus, they defined two important parameters, the *shape* and *scale* of an image. They attempted to predict the number of views each image obtained over a period of 30 days.For each image, the views it receives at the end of 30 days is called the scale. Once the 30 day sequence is divided by the scale, a sequence varying between 0-1 is obtained. This sequence is called the shape. Accordingly they stated

$$seq_{scale} = max[v_1, v_2...v_n] \qquad (2)$$

$$seq_{shape} = [v_1/seq_{scale}, v_2/seq_{scale}...v_n/seq_{scale}] \qquad (3)$$

Finally predicting the scale and the sequence individually, the entire sequence is obtained as

$$seq = seq_{scale} * seq_{shape} \qquad (4)$$

3   Implemented Methodology

In the present work, a strategy similar to that suggested in [2] has been followed. The first and foremost thing to do was data pre-processing. Since the dataset

contained real world data, there were a lot of missing values. Data interpolation had to be done. The features that contained string were converted into embeddings. Following this, the engagement scores of the images were analysed by plotting the values. Figure 1(a) shows one such example. On plotting the engagement values of different images, it was observed that two different images may have the same engagement dynamics but the views each image obtains may be very different, as was pointed out in [2]. This can be observed on comparing Figure 1(a) and Figure 1(b).

(a) No of Views of Image-86 over 30 days

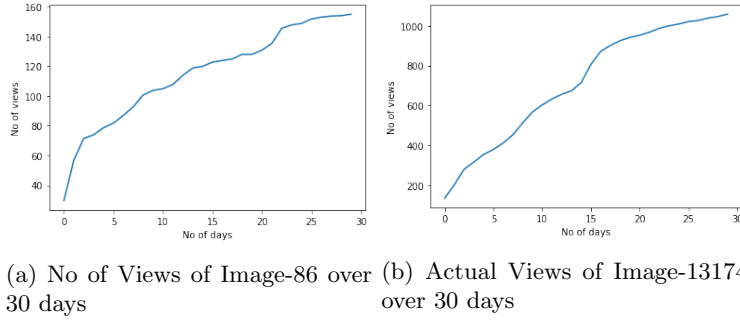(b) Actual Views of Image-13174 over 30 days

Figure 1: Views of two different images having similar dynamics

Therefore, this makes it difficult to compare different images because of their different scales. So to bring them all to the same scale, the engagement sequence of each image was divided by it's maximum number of views, that is the views obtained by each image till the 30th Day. The maximum views that each image obtains, is referred to as the *scale* of that image. Thereby on dividing the views obtained by each image over 30 days by their respective scale, a normalised set of views is obtained for each image. Figure 2(a) shows an image with it's actual views while Figure 2(b) shows the views for the same image scaled between 0-1.

(a) No of Views of Image-86 over 30 days
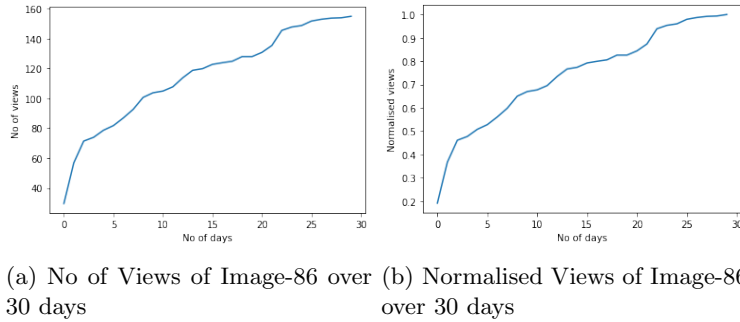
(b) Normalised Views of Image-86 over 30 days

Figure 2: Actual Views and Normalised Views of the Same Image

Once the views for all the images were scaled down to 0-1, the engagement sequence of the images along were plotted. Figure 3(a) shows what it looks like

with the views for all the images to be scaled down to 0-1. Therefore, it can be observed that no such proper number of clusters can be visualised from it. However, on plotting 5 random images as in Figure 3(b), it can be seen that the shapes of some images are similar, hinting to the existence of clusters.
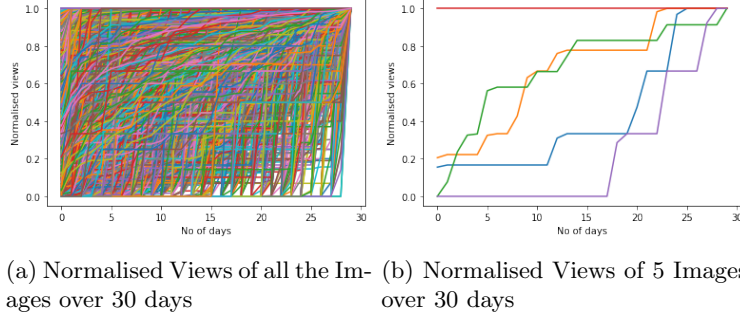

(a) Normalised Views of all the Images over 30 days

(b) Normalised Views of 5 Images over 30 days

Figure 3

3.1 Shape Allotment

In [2],the authors implemented K-means for the purpose of clustering the images based on their normalised views. They observed that for **k=50**, optimal results were obtained. Therefore, in the present work, the following methods were first explored to find out and confirm the optimal number of clusters that could be obtained using K-means.

3.1.1 Elbow Method

The dataset was clustered by varying the number of clusters from 1-80 (since the optimal suggested was 50) in an attempt to verify the optimal number of clusters. The Elbow Method shows the total WSS(Within Cluster Sum of Squares) for each cluster number. A lower WSS hints at proper clustering of the data. Once an *elbow like* shape is observed, the cluster number corresponding to that point gives the optimal number of clusters. On calculating the total WSS , a graph of WSS vs No of clusters was plotted as shown in Figure 4.

Therefore, it can be observed that no proper elbow is visible especially not at k=50 but that can be expected as Elbow Method is a Naive Approach.
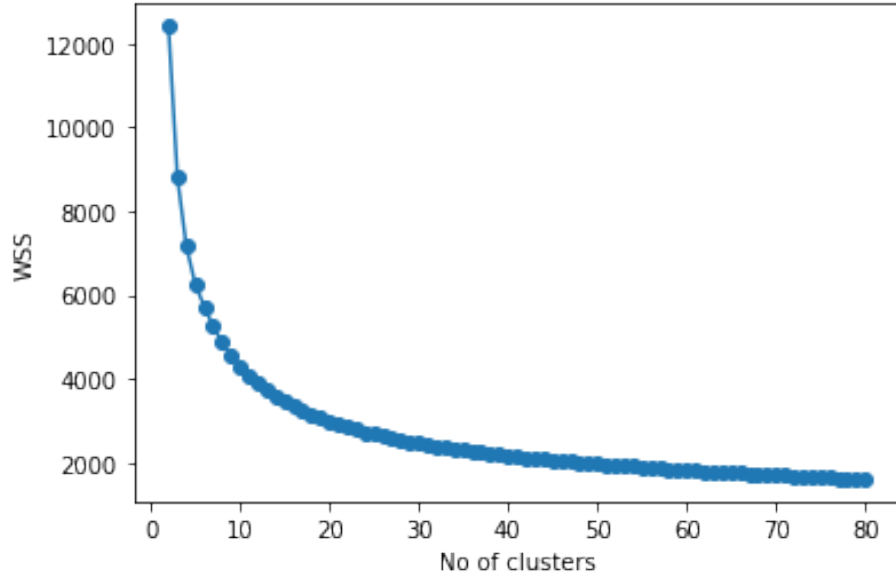
4

Figure 4: Elbow Method Analysis

3.1.2 Silhouette Score

The next method attempts to find the optimal number of k by calculating the Silhouette score for each cluster number. A Silhouette score displays a measure of how close each point in one cluster is to points in the neighboring clusters. By varying the number of clusters from 2-80 (because Silhouette score calculation demands at least 2 clusters), the Silhouette score for each cluster number was plotted as shown in Figure 5.

A higher Silhouette score is desirable because it shows that the data points within a cluster are more similar to the points in the cluster and dissimilar to the points in other clusters. However, it may not always give a proper result (which is k=2 here) because it may simply imply that the clusters have been formed in such a way that the clusters are away from each other but aren't proper. Furthermore, it is obvious that k=2 cannot be possible, since from Figure 3(b) it can be seen that at least 3 different clusters exist.
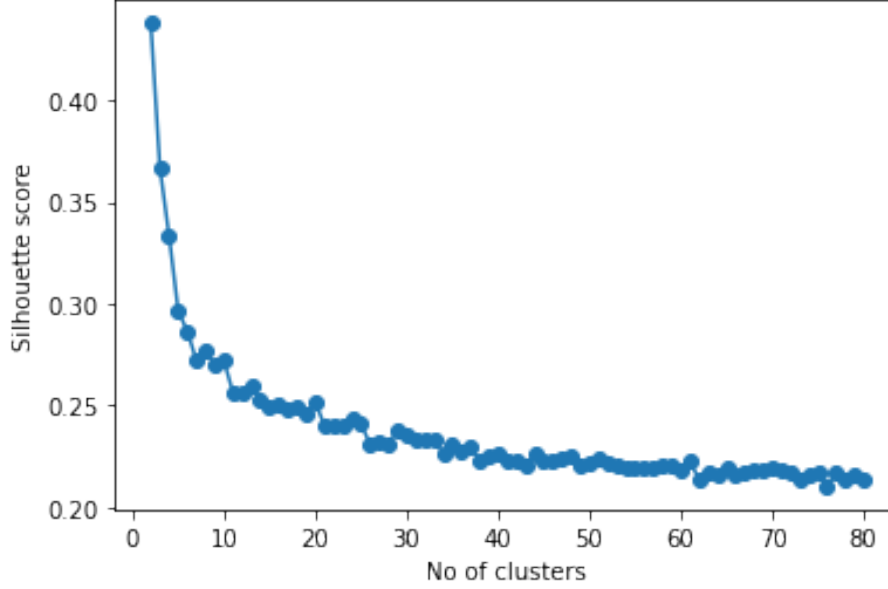
5

Figure 5: Silhouette Score

3.2 Shape Allotment Analysis

In **Section 3.1.1** and **Section 3.1.2**, it is observed that no optimal k was found, neither k=50. Furthermore, a Hierarchical clustering (Mean Shift Clustering) was also tried out to test if it performs any better than K-means. The data was clustered using K-means with k=50 as suggested by [2]. Whereas, with the *bandwidth as 0.53* for Mean Shift clustering, **49 clusters** were obtained. A Random Forest classifier (RNF) was chosen as it uses ensemble method with several decision trees and thus gives the best of classification results. The optimal parameters for the RNF were obtained using a Grid-Search method. Thus, with the same set of parameters, a comparison between both the clustering methods was done. All the social features related to the user and posts were provided as input into the RNF.

Accuracy on using K-means Clustering = 20%

Accuracy on using Mean Shift Clustering = 62%

With even the number of clusters almost equal, Mean Shift clustering gave better results than K-means,thus making it our preferred choice. Once the clustering was done, the cluster centroids were treated as the *shape prototypes*. Each image was then classified into their clusters using RNF.

6

3.3 Scale Prediction

Once the shape allotment was done, the scale of each image had to be predicted. This was done using a SVR (Support Vector Regressor). Social features related to the post were fed into the SVR one at a time. Then the Spearman's Rank Correlation of each feature with the maximum views , i.e the scale was recorded. Even the visual features were extracted from each image exploiting three state of the art Convolutional Neural Networks (CNN).(Hybridnet [4], DeepSentiBank [5] and GoogleNet [6]). Only the last two layers of activations before the softmax were used to extract the features. However, they contributed very little to predicting the scale of the image, as had been predicted in [1]. Finally, a suitable combination of the features was used so as to get a decent rank correlation. Once the necessary features were selected, the scale of the images were predicted using the SVR. Table 1 shows some of the highest Rank Correlations obtained by considering some individual features and groups of features.

Features	Rank Correaltion (Spearman)
MeanViews	0.7749
Counts	0.556
GroupsCount	0.562
MeanViews+GroupsCount	0.7448
MeanViews+Contacts	0.7465
MeanViews+Contacts+GroupsCount	0.7469
MeanViews+Contacts+GroupsCount +GroupsAvgMemb+GroupsAvgPic	0.744

Table 1: Rank Correlation of Features

4 Results

The training and testing of the images was done based on 9:1 train, test split. This was repeated for 10 runs and the final result was obtained by averaging the results of each run. After obtaining the scale and shape individually, on multiplying the *scale* with the *shape* of each image, a 30 day predicted engagement sequence of the images was obtained. However, a problem with predicting the scale was that, it was unbounded. Therefore, the model could end up predicting both very high as well as very low values. Thus, evaluating the results using an RMSE could lead to very high errors because of some outliers. Therefore, to make a more meaningful evaluation, 25% of the highest and and the lowest predicted values were trimmed off.The final results were evaluated using the 25% trimmed RMSE (interquartile mean) and the Median RMSE. Some of the

best results obtained using different considered features and the corresponding rank correlation as well as the final errors, are presented in Table 2.

Features	RankCorrelation	tRMSE 0.25	tRMSE MED
MeanViews+Contacts+GroupsCount+ GroupsAvgMembers+GroupsAvgPictures+ NumGroups+AvgGroupsMemb+AvgGroupsPhotos	0.753	12.2	8.5
Contacts+PhotoCount+MeanViews+ NumGroups+AvgGroupMembers+AvgGroupPhotos	0.761	12.06	8.25

Table 2: Results obtained by considering some features

5   Conclusion

In this work, following a method similar to the one suggested in [2], the number of views an image obtains over a period of 30 days was predicted. This is much more challenging than predicting the total amount of views an image receives at the end of a period since this involves considering the engagement dynamics of an image. The shape allocation of images is something that could be worked on. The method followed in the current work requires us to cluster the data across 30 columns. Therefore, each data point is a 30 dimensional vector. Thus, making it clustering in higher dimensions and clustering on such high dimension affects results. The problem is, using the current approach dimensional reduction is not possible because all the 30 cluster centroids are required to define the shape prototype of the plots. Therefore, some other method wherein fewer number of features to cluster the plots can be used and accordingly their centroids can be calculated to get the shape prototypes we desire.

6   References

1.Khosla, Aditya, Atish Das Sarma, and Raffay Hamid. "What makes an image popular?." Proceedings of the 23rd international conference on World wide web.pp 867-876 ACM, 2014.
2. Ortis, Alessandro, Giovanni Maria Farinella, and Sebastiano Battiato. "Prediction of Social Image Popularity Dynamics." International Conference on Image Analysis and Processing. Springer, Cham, 2019, pp 572-582.
3. A. Ortis, G. M. Farinella and S. Battiato, "Predicting Social Image Popularity Dynamics at Time Zero" in IEEE Access, 2019, 7, pp. 1–15.
4. Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., Oliva, A.: Learning deep

features for scene recognition using places database. In: Advances in neural information processing systems. pp. 487–495 (2014)

5. Borth, D., Ji, R., Chen, T., Breuel, T., Chang, S.F.: Large-scale visual sentiment ontology and detectors using adjective noun pairs. In: Proceedings of the 21st ACM international conference on Multimedia. pp. 223–232. ACM (2013)

6. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D.,Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: In proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015)

7. Valafar, M., Rejaie, R., Willinger, W.: Beyond friendship graphs: a study of user interactions in flickr. In: Proceedings of the 2nd ACM workshop on Online social networks. pp. 25–30. ACM (2009)